1962

Maximum likelihood estimation of distribution parameters from incomplete data

Edwin Joseph Hughes
Iowa State University

HUGHES, Edwin Joseph, 1928–
MAXIMUM LIKELIHOOD ESTIMATION OF DISTRIBUTION PARAMETERS FROM INCOMPLETE DATA.

Iowa State University of Science and Technology
Ph.D., 1962
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan

MAXIMUM LIKELIHOOD ESTIMATION OF DISTRIBUTION PARAMETERS FROM INCOMPLETE DATA

by

Edwin Joseph Hughes

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1962

TABLE OF CONTENTS

INTRODUCTION

One of the fundamental problems of mathematical statistices involves the estimation of parameters which characterize a population from the information contained in a sample. In virtue of the fact that many distributions (populations) can be characterized by their moments, one general method of estimation involves the solution for the parameter values of estimating equations obtained by equating population moments to sample moments. Another general method of estimation involves the solution for parameter values of estimating equations obtained by maximizing, with respect to the parameters, the probability or likelihood function of the sample. Estimates obtained by the method of moments are usually easy to calculate, but they may be inefficient, in that more precise estimates can be obtained from the same data. Estimates from the method of maximum likelihood are, for some populations, the same as those from the method of moments. Maximum likelihood estimates are sometimes difficult to calculate and are sometimes biased. However, under very general regularity conditions (cf. Cramer (1949), pages 500, 504) they are asymptotically unbiased and of maximum precision. That is, in large samples the maximum likelihood estimates are unbiased and fully efficient. Estimates which are unbiased for small as well as large samples and are of maximum precision can be obtained by a third method of estimation if the population parameters have complete sufficient statistics. Since this property is the exception rather than the rule, the method of complete sufficient statistics is not of general application. Because of its general applicability and desirable large sample properties, the method of maximum likelihood will be used in our discussion of estimation from incomplete data.

When the parameters which characterize a population or distribution are to be estimated from a sample, it is required that certain (moment, likelihood) relationships be established between observations in the sample and paramters in the population. These relationships will exist only when, in the selection of the sample, each observation has the same distribution as the population. That is, these relationships will exist only when we take a random sample.

Now it sometimes happens that a variate value occurring in a population is not observable in a sample. For example, in a Poisson population of number of attacks of subjects exposed to a disease, the zero value is not observable in a sample, since exposed subjects with zero attacks are indistinguishable from unexposed subjects. In this case a random sample can be obtained only from the truncated or conditional Poisson distribution with zero-class missing. In other cases, for convenience, perhaps (say, in avoiding measurement of extreme values of the variate), a random sample is taken from a deliberately truncated distribution. In such examples as these we say that, with respect to the untruncated distribution, we have incomplete truncated data. The estimation of the parameters of truncated distributions, or, equivalently, the estimation of distribution parameters from incomplete truncated data, is part of the present study. One would like to avoid having a separate estimating procedure for each possible type of truncation.

It also happens that sometimes individual variate values occurring in a population are observed only as grouped together in a sample. Thus, if sample values are obtained by the use of some measuring device which has a limited range, then all population variate values greater than the maximum measurable value--when observed in a sample--are necessarily grouped together. The sample provides information only on the total number or count of these grouped values. In this case a random sample can be obtained only from a grouped or censored distribution. We say that, with respect to the ungrouped or uncensored distribution, we have incomplete censored data. The estimation of the parameters of censored distributions, or, equivalently, the estimation of distribution parameters from incomplete censored data, is part of the present study. Again, one would like to avoid having a separate estimation procedure for each possible type of censorship.

The early work on the estimation of distribution parameters from incomplete data, and much of the later work, is characterized by: 1) methods developed are specific to one particular distribution, 2) only special types of incompleteness are considered, and 3) special aid tables are required of functions other than distribution areas and ordinates, and often depending upon the nature of the incompleteness.

It is the purpose of this study to describe two procedures for the estimation of distribution parameters of a wide range of distributions, both discrete and continuous, from incomplete data of a very general type, requiring in most cases only tables of distribution areas and ordinates.

Both procedures to be described are for maximum likelihood estimation. The first procedure arises from the observation that the only complicating factor introduced in maximum likelihood estimating equations by the incompleteness of the data--namely, the derivative with respect to the parameter of the cumulative distribution function--can, for many common distributions, both discrete and continuous, be expressed in terms of the distribution area and ordinate functions. In such cases the maximum likelihood equations, seldom giving explicit expressions for the parameters, can be solved by simple iterative, interpolative procedures with the use only of tables of distribution areas and ordinates.

The second estimation procedure is that described by Hartley (1958) for discrete distributions and extended by Krane (1957) to continuous distributions. Hartley's method is extended to continuous distributions here in a way different from that described by Krane, and a unified treatment is given for the two cases. In addition, a theorem is proved establishing sufficient conditions for the convergence of the required iterative procedure. Hartley's method arises from the observation that, by introducing "pseudo-frequencies" for variate values in the subsets of truncation and censorship (under a system of "proportional allocation"), the maximum likelihood estimating equations from incomplete data can be put into the form of maximum likelihood equations from complete data. The latter equations are solved by the usual complete data methods. The pseudo-frequencies are calculated from proportional allocation equations requiring the use only of tables of distribution areas and ordinates. Initial estimates of the parameters are used to obtain pseudo-frequencies, and these are substituted in the maximum likelihood estimating equations for complete data to obtain improved estimates. The process is repeated until convergence, when the maximum likelihood estimates of the parameters are obtained.

INCOMPLETE DATA

Introduction

We shall define incomplete data as a combination of incomplete truncated data and incomplete censored data. The latter two concepts are defined, respectively, in terms of a truncated distribution and a censored distribution. These distributions are derived from a given or "parent" distribution. A random sample from a truncated distribution is said to constitute incomplete truncated data (incomplete with respect to the untruncated or parent distribution). A random sample from a censored distribution is said to constitute incomplete censored data (incomplete with respect to the uncensored or parent distribution). All these concepts are considered in detail below. The discrete case is considered separately from the continuous case. It should be noted that, with respect to the subsets of truncation T and censorship C, the continuous case is not treated in the full generality of the discrete case.

Discrete Case

Parent distribution

Let Y be a random variable having a discrete distribution with parameter vector θ:

$$\Pr(Y = y; \theta) = p(y; \theta), \qquad y \in R = \{0, 1, 2, \cdot \cdot \cdot\}, \tag{1}$$

$$\Pr(Y \leq y;\ \theta) = P(y;\ \theta), \qquad y \epsilon R. \tag{2}$$

The probability measure P defined on the set of variate values R is said to define the parent distribution. The term "parent distribution" is used to distinguish a given distribution from distributions derived from it, as, for example, conditional distributions.

Truncated (conditional) distribution

Given the probability measure P defined on the set R by Equation 2, and an arbitrary subset $T \subsetneqq R$, let Y' be a random variable having a distribution with parameter vector θ such that

$$\Pr(Y' \epsilon T;\ \theta) = 0, \qquad T \subsetneqq R, \tag{3}$$

$$\Pr(Y' = y';\ \theta) = \frac{p(y';\ \theta)}{[1 - \sum\limits_{y \epsilon T} p(y;\ \theta)]}, \quad y' \epsilon R-T. \tag{4}$$

The random variable Y' is said to have a truncated discrete distribution. In the set of variate values R, the subset of truncation is the set T.

Censored (grouped) distribution

Given the probability measure P defined on the set R by Equation 2, and an arbitrary subset $C \subsetneqq R$ consisting of at least two elements of R, let Y' be a random variable having a distribution with parameter vector θ such that

$$\Pr(Y' \epsilon C;\ \theta) = \sum_{y \epsilon C} p(y;\ \theta), \qquad C \subsetneqq R, \tag{5}$$

$$\Pr(Y' = y';\ \theta) = p(y';\ \theta), \qquad y' \epsilon R-C, \tag{6}$$

but $\Pr(Y' = y';\ \theta)$ is undefined for $y' \epsilon C$. The random variable Y' is said to have a censored discrete distribution. In the set of variate values R, the subset of censorship is the set C.

Random sample

When we say that a sample of n observations from a given distribution is a random sample, we mean that, for each observation X in the sample, the distribution of X is the same as the given distribution.

Complete data

A random sample from the parent distribution is said to constitute complete data. In the case of complete data, a sample of n observations consists of n values from R. For each observation X,

$$\Pr(X = x;\ \theta) = p(x;\ \theta), \qquad x \epsilon R = \{0, 1, 2, \cdot \cdot \cdot\}, \tag{7}$$

$$\Pr(X \leq x;\ \theta) = P(x;\ \theta), \qquad x \epsilon R. \tag{8}$$

Incomplete truncated data

A random sample from a truncated distribution is said to constitute incomplete truncated data (incomplete with respect to the untruncated or parent distribution). In the case of incomplete truncated data, a sample of n observations consists of n values from $R' = R-T$. For each observation X,

$$\Pr(X\epsilon T;\ \Theta) = 0, \qquad T \subsetneqq R, \tag{9}$$

$$\Pr(X = x;\ \Theta) = \frac{p(x;\ \Theta)}{[1 - \sum\limits_{y\epsilon T} p(y;\ \Theta)]} \qquad x\epsilon R' = R-T. \tag{10}$$

<u>Incomplete censored data</u>

A random sample from a censored distribution is said to constitute incomplete censored data (incomplete with respect to the uncensored or parent distribution). In the case of incomplete censored data, a sample of n observations consists of, say, m counts from C (where m is a random variable, $0 \leq m \leq n$) and $n' = n-m$ values from $R' = R-C$. For each observation X,

$$\Pr(X\epsilon C;\ \Theta) = \sum_{y\epsilon C} p(y;\ \Theta), \qquad C \subsetneqq R, \tag{11}$$

$$\Pr(X = x;\ \Theta) = p(x;\ \Theta), \qquad x\epsilon R' = R-C. \tag{12}$$

<u>Incomplete data (incomplete truncated and censored data)</u>

A random sample from a truncated and censored distribution is said to constitute incomplete (truncated and censored) data. In the case of incomplete truncated and censored data, a sample of n observations consists of, say, m counts from C (where m is a random variable, $0 \leq m \leq n$) and $n' = n-m$ values from $R' = R-T-C$. For each observation X,

$$\Pr(X\epsilon T;\ \Theta) = 0, \qquad T \subsetneqq R, \tag{13}$$

$$\Pr(X\epsilon C;\ \Theta) = \frac{\sum\limits_{y\epsilon C} p(y;\ \Theta)}{[1 - \sum\limits_{y\epsilon T} p(y;\ \Theta)]}, \qquad C \subsetneqq R-T, \tag{14}$$

$$\Pr(X = x;\ \theta) = \frac{p(x;\ \theta)}{[1 - \sum_{y \in T} p(y;\ \theta)]},\quad x \in R' = R-T-C. \tag{15}$$

<u>Incomplete data, general case (incomplete truncated and censored data with a finite number of subsets of censorship)</u>

Let $T \subsetneqq R$ and $C_1, C_2, \cdots, C_Q \subsetneqq R-T$ and non-intersecting. A sample in which each observation X has the distribution

$$\Pr(X \in T;\ \theta) = 0, \qquad T \subsetneqq R, \tag{16}$$

$$\Pr(X \in C_q;\ \theta) = \frac{\sum_{y \in C_q} p(y;\ \theta)}{[1 - \sum_{y \in T} p(y;\ \theta)]} \qquad C_q \subsetneqq R-T,\quad q = 1, 2, \cdots, Q, \tag{17}$$

$$\Pr(X=x;\ \theta) = \frac{p(x;\ \theta)}{[1 - \sum_{y \in T} p(y;\ \theta)]},\quad x \in R' = R-T-\Sigma C_q \tag{18}$$

is said to constitute incomplete (truncated and censored) data, general case. Here a sample of n observations consists of, say, m_q counts from C_q, $q = 1, 2, \cdots, Q$ (where the m_q are random variables, $0 \leq m_q, \Sigma m_q \leq n$), and $n' = n - \Sigma m_q$ values from $R' = R-T-\Sigma C_q$.

Continuous Case

<u>Parent distribution</u>

Let Y be a random variable having a continuous distribution with parameter vector θ:

$$\Pr(Y\epsilon(y,\ y+dy);\ \theta) = p(y;\ \theta)dy, \qquad y\epsilon R = (-\infty,\ \infty), \tag{19}$$

$$\Pr(Y \leq y;\ \theta) = P(y;\ \theta), \qquad y\epsilon R. \tag{20}$$

The probability measure P defined on the set of variate values R is said to define the parent distribution.

Incomplete data, general case (incomplete truncated and censored data with a finite number of subsets of censorship)

Let $T \subsetneqq R$, $T = \Sigma I_j$, a finite sum of non-degenerate, non-overlapping intervals; let $C_1, C_2, \cdot\cdot\cdot, C_Q \subsetneqq R-T$, each C_q, $q = 1, 2, \cdot\cdot\cdot, Q$, a finite sum of non-degenerate, non-overlapping intervals, $C_q = \sum_j I_{qj}$, and the C_q non-intersecting. A sample in which each observation X has the distribution

$$\Pr(X\epsilon T;\ \theta) = 0, \qquad T = \Sigma I_j \subsetneqq R, \tag{21}$$

$$\Pr(X\epsilon C_q;\ \theta) = \frac{\int_{C_q} p(y;\ \theta)dy}{[1 - \int_T p(y;\ \theta)dy]}, \quad C_q = \sum_j I_{qj} \subsetneqq R-T, \quad q = 1, 2, \cdot\cdot\cdot, Q, \tag{22}$$

$$\Pr(X\epsilon(x,x+dx);\ \theta) = \frac{p(x;\ \theta)dx}{[1 - \int_T p(y;\ \theta)dy]}, \quad x\epsilon R' = R-T-\Sigma C_q \tag{23}$$

is said to constitute incomplete (truncated and censored) data, general case. Here a sample of n observations consists of, say, m_q counts from C_q, $q = 1, 2, \cdot\cdot\cdot, Q$ (where the m_q are random variables, $0 \leq m_q$, $\Sigma m_q \leq n$), and $n' = n - \Sigma m_q$ values from $R' = R - T - \Sigma C_q$.

REVIEW OF LITERATURE

It was observed above, in the introduction, that the early work on the estimation of distribution parameters from incomplete data, and much of the later work, is characterized by: (1) methods developed are specific to one particular distribution, (2) only special types of incompleteness are considered, and (3) special aid tables are required of functions other than distribution areas and ordinates, often depending upon the nature of the incompleteness. This is easily observed in a survey of the literature.

Pearson and Lee (1908) consider the normal distribution truncated in a single tail, and use the method of moments (taking moments about the point of truncation) to obtain an estimate of the standardized point of truncation, and, from this estimate, estimates of the distribution parameters μ, σ. Two special aid tables required in the estimation procedure are given in the paper.

Fisher (1931) considers the same problem as Pearson and Lee (1908), namely the normal distribution truncated in a single tail, and shows that the moments estimates for this problem are identical to the maximum likelihood estimates. Fisher provides special aid tables for his version of the estimation procedure and gives asymptotic expressions for the variances and covariance of the estimates.

Tippett (1932) studies the Poisson distribution censored in the upper tail, and uses the method of maximum likelihood to obtain an estimate of the Poisson parameter. Nomograms and a special aid table are provided for use in the estimation procedure. An expression is given for the asymptotic variance of the estimate.

Hald (1949) studies the normal distribution censored or truncated in a single tail, and derives similar maximum likelihood estimating equations for the two cases. Observations are transformed so that the point of truncation is the origin. As in the Pearson-Lee and Fisher methods, one obtains from the data an estimate of the standardized point of truncation, and, from this estimate, estimates of the distribution parameters μ, σ. Hald gives the special aid tables required for the application of his method and the estimation of the asymptotic variances and covariance of the estimates.

Gjeddebaek (1949) considers the completely censored normal distribution, and uses the method of maximum likelihood to obtain estimating equations for the normal parameters. He provides special aid tables to assist in the solution of the estimating equations. Estimates of the asymptotic variances and covariance of the estimates are given.

Finney (1949) considers the binomial distribution truncated in a single tail or in both tails, and applies a method of "successive approximations" (system of scoring) to solve the maximum likelihood estimating equation for an estimate of the binomial parameter p. An estimate of the variance of the estimate is obtained as part of the estimating procedure. Finney provides special aid tables to facilitate the use of his method.

Cohen (1949) observes that the tabulated functions in the special aid tables of Pearson-Lee and Fisher can be expressed in terms of normal areas and ordinates, and, as a consequence, the estimation procedure can be carried out without the use of special aid tables. Cohen (1950) studies the normal distribution, and gives a unified treatment of maximum likelihood estimation from the following types of incomplete data: truncated in a single tail or in both tails, censored in a single tail or in both tails separately, censored in both tails combined.

Gupta (1952) considers the normal distribution and his so-called "Type II" censorship, where the m smallest (greatest) observations of a sample of size n are unmeasured. ("Type I" censorship is that where all observations smaller (greater) than a fixed point of censorship are unmeasured.) He uses the method of maximum likelihood and obtains results similar to those of Hald (1949), with the smallest (greatest) measured observation replacing the fixed point of censorship in all formulas. In addition, for problems involving small samples, Gupta uses the method of least squares to find best linear unbiased estimates of the parameters of a normal distribution

Type II censored in a single tail, and gives tables of the coefficients for the estimates and tables for estimating the variances and covariance of the estimates.

David and Johnson (1952) consider the Poisson distribution with zero-class missing, and use both the method of moments and the method of maximum likelihood to derive estimates of the Poisson parameter. It is shown that the easily computed moments estimate is never less than 70% efficient. A special aid table is given to facilitate the application of the maximum likelihood estimating procedure. In addition, the authors consider estimation of the two parameters of the negative binomial distribution with zero-class missing. They give the maximum likelihood estimates, and moments estimates involving the first three sample moments. The moments estimates are shown to be very inefficient, due primarily to the use of the third sample moment. An iterative procedure is suggested for the solution of the maximum likelihood equations, and a special aid table is provided to assist in the task.

Des Raj (1953) studies the Pearson Type III distribution expressed in terms of the population mean μ, standard deviation σ, and third standard moment α_3. The types of incompleteness considered are: truncation in a single tail or in both tails, censorship in a single tail or in both tails separately, censorship in both tails combined. Both the method of maximum likelihood and the method of moments are used to derive estimating equations, and only tables of the Pearson Type III areas and ordinates are required for their solution. Most results require that α_3 be known.

Moore (1954) uses the method he developed earlier, Moore (1952), exploiting an identity satisfied by the Poisson parameter, to obtain an easily computed, "almost unbiased" estimate for the parameter λ of a Poisson distribution truncated in the lower tail, or in the upper tail, or in a middle interval. Asymptotic expressions are derived for the variance of the estimates in the cases: truncation in the lower tail and truncation in the upper tail. The estimates are recommended as easily obtained estimates or as starting values for maximum likelihood solutions. Moore applies the same method to estimate the parameter p of a binomial distribution with zero-class missing.

Cohen (1954) considers the Poisson distribution, and gives a unified treatment of maximum likelihood estimation from the following types of incomplete data: truncated in a single tail or in both tails, censored in a single tail or in both tails separately, censored in both tails combined. By virtue of the fact that the derivative with respect to the Poisson parameter of the cumulative distribution function can be expressed in terms of the ordinate function, only tables of Poisson areas and ordinates are required in the estimation procedure. Estimates of the asymptotic variance of the estimate are derived for the various cases of incompleteness, and are easily obtained without the use of special aid tables.

Rider (1953) considers the Poisson distribution truncated in the lower tail, and, introducing "pseudo-frequencies" for the missing variate values, uses the method of moments on a "completed sample" to obtain an easily computed estimate of the Poisson parameter in terms of the first two sample moments. Rider's estimate reduces to the moments estimate of David and Johnson (1952) for the special case of the zero-class missing. Rider (1955) uses the same method for an estimate of the parameter p of the binomial distribution truncated in a single tail, and for estimates of the two parameters of the negative binomial distribution with zero-class missing. The negative binomial estimates are expressed in terms of the first three sample moments, are the same as those given by David and Johnson (1952), and, as was shown by David and Johnson, are very inefficient.

Sampford (1955) studies the negative binomial distribution with zero-class missing, and uses the method of moments to obtain estimating equations for the two distribution parameters involving only the first two sample moments. In contrast to the moments estimates given by David and Johnson (1952) for this problem, involving the first three sample moments and having low efficiency, the estimates obtained here have an efficiency of 80% or better for "all but the most unfavorable combinations of the parameters." But the two estimating equations must be solved by trial and error, or iteratively. A special aid table is provided to assist in the solution. Asymptotic expressions are obtained for the variances and covariance of the estimates.

den Broeder (1955) considers the gamma or Pearson Type III distribution truncated in a single tail or censored in a single tail, and uses the method of maximum likelihood to derive an estimating equation for the scale parameter, when the other parameter is known, for each of the cases of incompleteness considered.

Deemer and Votaw (1955) consider the single parameter exponential distribution truncated or censored in the upper tail, and use the method of maximum likelihood to obtain an estimate of the scale parameter. For the case of censorship in the upper tail, an explicit expression for the estimate is obtained. A special aid table is provided for the case of truncation in the upper tail. In each case an asymptotic expression for the variance of the estimate is given.

Sarhan and Greenberg (1956, 1957) extend the small sample work of Gupta (1952), considering the normal distribution and the one and two parameter exponential distributions with Gupta's Type II censorship in a single tail or in both tails separately, and use the method of least squares to obtain best linear unbiased estimates of the distribution parameters. For each combination of distribution, parameter, type of incompleteness, and sample size, tables of coefficients for the appropriate estimate are given for samples of size ≤ 10. In addition, tables of variances and covariances of the estimates are given for the various types of incompleteness considered for samples of size ≤ 10.

Hartley (1958) develops a method, applicable to any discrete distribution, of reducing the problem of maximum likelihood estimation from incomplete data to that from complete data. The incompleteness of the data may be of an arbitrary type: truncation or censorship in a single tail, or in both tails, or in a central section, or combinations of any of these are but special cases. The reduction from incomplete to complete data is accomplished by the introduction of "pseudo-frequencies" (under a scheme of "proportional allocation") for classes that are missing (truncation) or grouped (censorship). An iterative procedure is given for the solution of the estimating equations in which no special aid tables are required. Asymptotic variances and covariances of estimates of the distribution parameters are approximated by certain first order divided differences involving derivatives of the log likelihood function.

Krane (1957) gives an extension of the method of Hartley (1958) to continuous distributions.

Tate and Goen (1958) consider the Poisson distribution truncated in the lower tail, and use the method of complete sufficient statistics to obtain a minimum variance unbiased estimator. In the special case where only the zero-class is missing, the estimate is expressed in terms of Stirling numbers of the second kind. Special aid tables of Stirling numbers of the second kind and of "generalized Stirling numbers" are provided to facilitate the evaluation of the estimates.

Clark and Williams (1960) consider arbitrary continuous distributions with arbitrary intervals of censorship, and develop a method of estimating the distribution parameters from relations obtained by equating certain sample statistics to their expected values. The statistics used are described as generalizations of the maximum likelihood estimator of the scale parameter of an exponential distribution censored in the upper tail, and consist of moments about zero of measured sample values augmented with weighted mid-point or end-point values, one from each interval of censorship. Estimates of variances and covariances are derived. The authors state it is intuitively clear that the statistics have roughly the same efficiency as estimators used in the classical method of moments. Two special aid tables are provided to assist in the evaluation of a term in the expected values common to all distributions.

METHOD 1

Introduction

The first procedure to be considered for maximum likelihood estimation of distribution parameters from incomplete data will be designated as Method 1. For incomplete data of a very general type from a given distribution, Method 1 involves the solution of maximum likelihood estimating equations by simple iterative, interpolative procedures with the use only of tables of the distribution areas and ordinates. We shall see that the method applies to many common distributions, both discrete and continuous.

Likelihood Function for Incomplete Data

We consider a parent distribution, unspecified as being discrete or continuous. Let Y be a random variable having a distribution with parameter vector θ:

(discrete case) $\Pr(Y = y;\ \theta) = p(y;\ \theta), \quad y \in R = \{1,\ 2,\ \cdots\},$ (24a)

(continuous case) $\Pr(Y \in (y, y+dy);\ \theta) = p(y;\ \theta)dy, \quad y \in R = (-\infty,\ \infty),$ (24b)

$$\Pr(Y \leq y;\ \theta) = P(y;\ \theta), \quad y \in R. \tag{25}$$

Incomplete data, general case, has already been defined (cf. Equations 16, 17, 18; 21, 22, 23) as follows. Let $T \subsetneqq R$ and $C_1, C_2, \cdots, C_Q \subsetneqq R-T$ and non-intersecting. (In the discrete case T and the C's are arbitrary subsets of R, where-as in the continuous case they must be finite sums of non-degenerate, non-overlapping intervals in R.) A random sample in which each observation X has the distribution

$$\Pr(X\epsilon T;\ \theta) = 0, \qquad T \subsetneqq R, \tag{26}$$

$$\Pr(X\epsilon C_q;\ \theta) = \frac{\Pr(Y\epsilon C_q;\ \theta)}{[1 - \Pr(Y\epsilon T;\ \theta)]}, \quad C_q \subsetneqq R-T, \quad q = 1, 2, \cdot\cdot\cdot, Q, \tag{27}$$

(discrete case)

$$\Pr(X = x;\ \theta) = \frac{p(x;\ \theta)}{[1 - \Pr(Y\epsilon T;\ \theta)]}, \quad x\epsilon R' = R-T-\Sigma\ C_q \tag{28a}$$

(continuous case)

$$\Pr(X\epsilon (x,x+dx);\ \theta) = \frac{p(x;\ \theta)\,dx}{[1 - \Pr(Y\epsilon T;\ \theta)]}, \quad x\epsilon R' \tag{28b}$$

is said to constitute incomplete data, general case. Here a sample of n observations consists of, say, m_q counts from C_q, $q = 1, 2, \cdot\cdot\cdot, Q$ (where the m_q are random variables, $0 \leq m_q$, $\Sigma\ m_q \leq n$), and $n' = n - \Sigma\ m_q$ values from $R' = R-T-\Sigma\ C_q$.

The likelihood function for the sample of size n is

$P = P(x_1, x_2, \cdot\cdot\cdot, x_n;\ \theta)$,

$$P = \text{const.} \prod_{q=1}^{Q} \left(\frac{\Pr(Y\epsilon C_q;\ \theta)}{[1 - \Pr(Y\epsilon T;\ \theta)]} \right)^{m_q} \prod_{i=1}^{n'} \left(\frac{p(x_i;\ \theta)}{[1 - \Pr(Y\epsilon T;\ \theta)]} \right)$$

$$= \text{const.}\ [1 - \Pr(Y\epsilon T;\ \theta)]^{-n} \prod_{q=1}^{Q} [\Pr(Y\epsilon C_q;\ \theta)]^{m_q} \prod_{i=1}^{n'} p(x_i;\ \theta). \tag{29}$$

Thus the likelihood function, apart from a constant factor, can be regarded as the product of three terms: one involving the subset of truncation T, one involving the subsets of censorship C_q, and one involving the remainder of the range $R' = R-T-\Sigma\, C_q$. If there is no truncation, then T is the null set and the first term reduces to unity. If there is no censorship, then each C_q is null, each m_q is zero, and the second term is replaced by unity. If there is complete censorship, so that R' is null and n' is zero (and Q is at least 2), then the third term is replaced by unity.

Estimating Equation for One Parameter Distributions

For simplicity we consider one parameter distributions with but a single subset of censorship. Thus the parameter vector θ is a scalar and $Q = 1$. Let $m = m_1$ and $C = C_1$. Then $n' = n-m$ and $R' = R-T-C$. In this case the likelihood function is $P = P(x_1, x_2, \cdot\cdot\cdot, x_n; \theta)$,

$$P = \text{const.}[\Pr(Y \epsilon R-T;\ \theta)]^{-n}[\Pr(Y \epsilon C;\ \theta)]^{m} \prod_{i=1}^{n'} p(x_i;\ \theta). \tag{30}$$

The log likelihood function $L = \ln P$ is

$$L = \text{const.} - n \ln \Pr(Y \epsilon R-T;\ \theta) + m \ln \Pr(Y \epsilon C;\ \theta) + \sum_{i=1}^{n'} \ln p(x_i;\ \theta). \tag{31}$$

And the maximum likelihood equation for estimating the parameter θ is

$$0 = \frac{dL}{d\theta} = -n \frac{d}{d\theta} \ln \Pr(Y \epsilon R-T;\ \theta) + m \frac{d}{d\theta} \ln \Pr(Y \epsilon C;\ \theta)$$

$$+ \sum_{i=1}^{n'} \frac{d}{d\theta} \ln p(x_i;\ \theta). \qquad (32)$$

Now for certain common discrete distributions (Poisson, binomial, negative binomial, geometric), and for certain common continuous distributions (normal, gamma, exponential, uniform) the derivatives

$$\frac{d}{d\theta} \ln \Pr(Y \epsilon R-T;\ \theta),\ \frac{d}{d\theta} \ln \Pr(Y \epsilon C;\ \theta),\ \frac{d}{d\theta} \ln p(y;\ \theta)$$

are easily expressed in terms of $P(y;\ \theta)$ and $p(y;\ \theta)$, so that, given a value θ' of θ, the quantity $dL(\theta')/d\theta$ can be evaluated from tables of distribution areas and ordinates alone (no special aid tables required). A maximum likelihood estimating procedure is, then, as follows:

(1) From an initial estimate of θ, find θ', θ'' such that $dL(\theta')/d\theta$ and $dL(\theta'')/d\theta$ are small in absolute value and on opposite sides of zero.

(2) Interpolate between θ' and θ'' for $\hat{\theta}$ such that $dL(\hat{\theta})/d\theta = 0$.

The two steps may be repeated with the interpolated value for θ replacing the initial estimate to obtain a more accurate estimate $\hat{\theta}$ of θ. This is Method 1 for one parameter distributions.

Discrete distributions

For discrete distributions probabilities can be expressed as sums of ordinate values,

$$\Pr(Y\epsilon R-T;\ \theta) = 1 - \sum_{y\epsilon T} p(y;\ \theta), \tag{33}$$

$$\Pr(Y\epsilon C;\ \theta) = \sum_{y\epsilon C} p(y;\ \theta), \tag{34}$$

and therefore the derivatives in the estimating equation are

$$d \ln \Pr(Y\epsilon R-T;\ \theta)/d\theta = [1 - \sum_{y\epsilon T} p(y;\ \theta)]^{-1}[- \sum_{y\epsilon T} dp(y;\ \theta/d\theta], \tag{35}$$

$$d \ln \Pr(Y\epsilon C;\ \theta)/d\theta = [\sum_{y\epsilon C} p(y;\ \theta)]^{-1}[\sum_{y\epsilon C} dp(y;\ \theta)/d\theta]. \tag{36}$$

Thus, in the discrete case, the maximum likelihood estimating equation for the parameter θ (cf. Equation 32) from incomplete (truncated and censored) data with a single subset of censorship is

$$\begin{aligned} 0 = dL/d\theta = {} & n[1 - \sum_{y\epsilon T} p(y;\ \theta)]^{-1}[\sum_{y\epsilon T} dp(y;\ \theta)/d\theta] \\ & + m[\sum_{y\epsilon C} p(y;\ \theta)]^{-1}[\sum_{y\epsilon C} dp(y;\ \theta)/d\theta] \\ & + \sum_{i=1}^{n'} d \ln p(x_i;\ \theta)/d\theta. \end{aligned} \tag{37}$$

It is clear from the above expression that, for a discrete distribution, $dL/d\theta$ can be expressed in terms of $P(y;\ \theta)$, $p(y;\ \theta)$ if $dp(y;\ \theta)/d\theta$ can be so expressed. Note, however, that if C (and similarly, T) consists of consecutive values of the variate, say, $C = \{c', c'+1, c'+2, \cdot\cdot\cdot, c''\}$, then

$$\Pr(Y\epsilon C;\ \theta) = \sum_{y\ C} p(y;\ \theta) = P(c'';\ \theta) - P(c'-1;\ \theta), \tag{38}$$

and

$$\frac{d}{d\theta} \ln \Pr(Y\epsilon C;\ \theta) = \frac{[dP(c'';\ \theta)/d\theta - dP(c'-1;\ \theta)/d\theta]}{[P(c'';\ \theta) - P(c'-1;\ \theta)]}. \tag{39}$$

In this case, with $T = \{t', t'+1, t'+2, \cdot\cdot\cdot, t''\}$, the estimating equation (cf. Equation 32) becomes

$$0 = \frac{dL}{d\theta} = n \frac{[dP(t'';\ \theta)/d\theta - dP(t'-1;\ \theta)/d\theta]}{[1 - P(t'';\ \theta) + P(t'-1;\ \theta)]}$$

$$+ m \frac{[dP(c'';\ \theta)/d\theta - dP(c'-1;\ \theta)/d\theta]}{[P(c'';\ \theta) - P(c'-1;\ \theta)]} + \sum_{i=1}^{n'} \frac{d}{d\theta} \ln p(x_i;\theta). \tag{40}$$

It is of interest, then, to have expressions for $dP(y;\ \theta)/d\theta$ as well as for $dp(y;\ \theta)/d\theta$ in terms of $P(y;\ \theta)$, $p(y;\ \theta)$. We give below examples of these expressions for certain discrete distributions.

<u>Poisson distribution</u>

$$p(y;\ \lambda) = e^{-\lambda}\lambda^y/y!, \qquad y\epsilon R = \{0, 1, 2, \cdot\cdot\cdot\}. \tag{41}$$

$$dp(y;\ \lambda)/d\lambda = (\lambda^{-1}y - 1)p(y;\ \lambda) \tag{42}$$

$$dP(y;\ \lambda)/d\lambda = -p(y;\ \lambda) \tag{43}$$

<u>Binomial distribution</u>

$$p(y;\ N,p) = \binom{N}{y} p^y(1 - p)^{N-y}, \qquad y\epsilon R = \{0, 1, 2, \cdot\cdot\cdot, N\}. \tag{44}$$

$$dp(y;\ N,p)/dp = (1-p)^{-1}N[(Np)^{-1}y - 1]p(y;\ N,p) \tag{45}$$

$$dP(y;\ N,p)/dp = -Np(y;\ N-1,p) \tag{46}$$

Negative binomial distribution

$$p(y;\ r,p) = \binom{y+r-1}{r-1}p^r(1-p)^y, \qquad y\epsilon R = \{0,\ 1,\ 2,\ \cdots\}. \tag{47}$$

$$\partial p(y;\ r,p)/\partial p = -p^{-1}r[r^{-1}(1-p)^{-1}py - 1]p(y;\ r,p) \tag{48}$$

$$\partial P(y;\ r,p)/\partial p = p^{-1}r[P(y;\ r,p) - P(y-1;\ r+1,p)] \tag{49}$$

$$\partial p(y;\ r,p)/\partial r = [\ln p + \sum_{z=0}^{y-1} 1/(r+z)]p(y;\ r,p) \tag{50}$$

$$\partial P(y;\ r,p)/\partial r = (\ln p)P(y;\ r,p) + \sum_{t=0}^{y}[\sum_{z=0}^{t-1} 1/(r+z)]p(t;\ r,p) \tag{51}$$

Geometric distribution

$$p(y;\ \theta) = \theta(1-\theta)^{y-1}, \qquad y\epsilon R = \{1,\ 2,\ 3,\ \cdots\}. \tag{52}$$

$$dp(y;\ \theta)/d\theta = -\theta^{-1}(1-\theta)^{-1}(\theta y - 1)p(y;\ \theta) \tag{53}$$

$$dP(y;\ \theta)/d\theta = (1-\theta)^{-1}y[1 - P(y;\ \theta)] \tag{54}$$

Continuous distributions

For simplicity in considering the continuous case, in addition to restricting ourselves to one parameter distributions with a single subset of censorship, we let both T and C consist of a "sum" of but a single interval. Denote T by (t', t") and C by (c', c"). We can write the probabilities occurring in the estimating equation (cf. Equation 32) as

$$Pr(Y\epsilon R-T;\ \theta) = 1 - P(t''\text{'}\ \theta) + P(t';\ \theta), \tag{55}$$

$$Pr(Y\epsilon C;\ \theta) = P(c'';\ \theta) - P(c';\ \theta). \tag{56}$$

Therefore, the derivatives in that equation are

$$\frac{d}{d\theta} \ln \Pr(Y\epsilon R-T;\ \theta) = \frac{[-dP(t'';\ \theta)/d\theta + dP(t';\ \theta)/d\theta]}{[1 - P(t'';\ \theta) + P(t';\ \theta)]}, \tag{57}$$

$$\frac{d}{d\theta} \ln \Pr(Y\epsilon C;\ \theta) = \frac{[dP(c'';\ \theta)/d\theta - dP(c';\ \theta)/d\theta]}{[P(c'';\ \theta) - P(c';\ \theta)]}. \tag{58}$$

Thus, in the continuous case, the maximum likelihood estimating equation for the parameter θ (cf. Equation 32) from incomplete (truncated and censored) data with a single subset of censorship is

$$0 = \frac{dL}{d\theta} = \frac{n[dP(t'';\ \theta)/d\theta - dP(t';\ \theta)/d\theta]}{[1 - P(t'';\ \theta) + P(t';\ \theta)]} + \frac{m[dP(c'';\ \theta)/d\theta - dP(c';\ \theta)/d\theta]}{[P(c'';\ \theta) - P(c';\ \theta)]} + \sum_{i=1}^{n'} \frac{d}{d\theta} \ln p(x_i;\ \theta). \tag{59}$$

It is clear from the above expression that, for a continuous distribution, $dL/d\theta$ can be expressed in terms of $P(y;\ \theta)$, $p(y;\ \theta)$ if $dP(y;\ \theta)/d\theta$ and $dp(y;\ \theta)/d\theta$ can be so expressed. Equation 59 should be compared with its discrete counterpart, Equation 40. Below we give examples of $dp(y;\ \theta)/d\theta$ and $dP(y;\ \theta)/d\theta$ for certain continuous distributions.

<u>Normal distribution</u>

$$p(y;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}[(y-\mu)/\sigma]^2}, \quad y\epsilon R = (-\infty,\ \infty). \tag{60}$$

$$\partial p(y;\ \mu,\sigma)/\partial\mu = \sigma^{-1}[(y-\mu)/\sigma]p(y;\ \mu,\sigma) \tag{61}$$

$$\partial P(y;\ \mu,\sigma)/\partial\mu = -p(y;\ \mu,\sigma) \tag{62}$$

$$\partial p(y;\ \mu,\sigma)/\partial\sigma = \sigma^{-1}([(y-\mu)/\sigma]^2-1)p(y;\ \mu,\sigma) \tag{63}$$

$$\partial P(y;\ \mu,\sigma)/\partial\sigma = -[(y-\mu)/\sigma]p(y;\ \mu,\sigma) \tag{64}$$

Gamma distribution

$$p(y;\ \alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\ y^{\alpha-1}\ e^{-y/\beta}, \qquad y\epsilon R = (0,\ \infty). \tag{65}$$

$$\partial p(y;\ \alpha,\beta)/\partial\alpha = [\ln(y/\beta) - \Gamma'(\alpha)/\Gamma(\alpha)]\ p(y;\ \alpha,\beta) \tag{66}$$

$$\partial P(y;\ \alpha,\beta)/\partial\alpha = \int_0^y (\ln t)p(t;\ \alpha,\beta)dt - [\ln\beta + \Gamma'(\alpha)/\Gamma(\alpha)]\ P(y;\ \alpha,\beta) \tag{67}$$

$$\partial p(y;\ \alpha,\beta)/\partial\beta = \beta^{-1}\alpha[(\alpha\beta)^{-1}y - 1]\ p(y;\ \alpha,\beta) \tag{68}$$

$$\partial P(y;\ \alpha,\beta)/\partial\beta = -\beta^{-1}y\ p(y;\ \alpha,\beta) \tag{69}$$

Exponential distribution

$$p(y;\ \theta) = (1/\theta)e^{-y/\theta}, \qquad y\epsilon R = (0,\ \infty). \tag{70}$$

$$dp(y;\ \theta)/d\theta = \theta^{-1}(\theta^{-1}y-1)p(y;\ \theta) \tag{71}$$

$$dP(y;\ \theta)/d\theta = -\theta^{-1}y\ p(y;\ \theta) \tag{72}$$

Uniform distribution

$$p(y;\ \theta) = 1/\theta, \qquad y\epsilon R = (0,\ \theta). \tag{73}$$

$$dp(y;\ \theta)/d\theta = -\theta^{-1}p(y;\ \theta) \tag{74}$$

$$dP(y;\ \theta)/d\theta = -\theta^{-1}yp(y;\ \theta), \qquad y < \theta. \tag{75}$$

Variance of the Estimate

The maximum likelihood estimate $\hat{\theta}$ of the distribution parameter θ is the solution of the maximum likelihood estimating equation. The asymptotic variance of $\hat{\theta}$ can be approximated by

$$\hat{V}(\hat{\theta}) = -[d^2L/d\theta^2]^{-1}_{\theta=\hat{\theta}}. \tag{76}$$

The second derivative $d^2L/d\theta^2$, and hence $\hat{V}(\hat{\theta})$, can be expressed in terms of $P(y;\ \theta)$, $p(y;\ \theta)$ in the same manner as the first derivative $dL/d\theta$ is so expressed. In the discrete case for arbitrary T, C, we differentiate Equation 37 with respect to θ to obtain an expression for $d^2L/d\theta^2$. In the discrete case where T, C consist of consecutive values of the variate, we differentiate Equation 40. In the continuous case, we differentiate Equation 59. For the continuous case we have

$$\begin{aligned}\frac{d^2L}{d\theta^2} &= n\,\frac{[d^2P(t'';\ \theta)/d\theta^2 - d^2P(t';\ \theta)/d\theta^2]}{[1 - P(t'';\ \theta) + P(t';\ \theta)]} \\ &+ n\,\frac{[dP(t'';\ \theta)/d\theta - dP(t';\ \theta)/d\theta]^2}{[1 - P(t'';\ \theta) + P(t';\ \theta)]^2} \\ &+ m\,\frac{[d^2P(c'';\ \theta)/d\theta^2 - d^2P(c';\ \theta)/d\theta^2]}{[P(c'';\ \theta) - P(c';\ \theta)]} \\ &- m\,\frac{[dP(c'';\ \theta)/d\theta - dP(c';\ \theta)/d\theta]^2}{[P(c'';\ \theta) - P(c';\ \theta)]^2} + \sum_{i=1}^{n'} \frac{d^2}{d\theta^2} \ln p(x_i;\ \theta).\end{aligned} \tag{77}$$

For Two Parameter Distributions

In the definition of the parent distribution (Equations 24a, 24b, 25) we have two coordinates for the parameter vector $\theta = (\theta_1, \theta_2)$. The maximum likelihood estimating equations are given by

$$0 = \partial L/\partial\theta_1 = L_1(\theta_1, \theta_2), \tag{78}$$

$$0 = \partial L/\partial\theta_2 = L_2(\theta_1, \theta_2) \tag{79}$$

(cf. Equations 32; 37, 40; 59). An iterative, interpolative procedure to solve the two equations simultaneously for the maximum likelihood estimates $(\hat{\theta}_1, \hat{\theta}_2)$ from starting values (approximate solutions) $({}_0\theta_1, {}_0\theta_2)$ is as follows. Let

${}_1\theta_2$ be the solution of $L_2({}_0\theta_1, \theta_2) = 0$,
${}_1\theta_1$ be the solution of $L_1(\theta_1, {}_1\theta_2) = 0$,
${}_2\theta_2$ be the solution of $L_2({}_1\theta_1, \theta_2) = 0$,
${}_2\theta_1$ be the solution of $L_1(\theta_1, {}_2\theta_2) = 0$.

Then

$({}_1\theta_1, {}_1\theta_2)$, $({}_2\theta_1, {}_2\theta_2)$ satisfy $L_1(\theta_1, \theta_2) = 0$;
$({}_0\theta_1, {}_1\theta_2)$, $({}_1\theta_1, {}_2\theta_2)$ satisfy $L_2(\theta_1, \theta_2) = 0$.

Approximate the two curves $L_1(\theta_1, \theta_2) = 0$ and $L_2(\theta_1, \theta_2) = 0$ in the vicinity of the solution $(\hat{\theta}_1, \hat{\theta}_2)$ by straight lines. We obtain, respectively,

$$(\theta_2 - {}_2\theta_2) = [({}_2\theta_2 - {}_1\theta_2)/({}_2\theta_1 - {}_1\theta_1)](\theta_1 - {}_2\theta_1), \tag{80}$$

$$(\theta_2 - {}_2\theta_2) = [({}_2\theta_2 - {}_1\theta_2)/({}_1\theta_1 - {}_0\theta_1)](\theta_1 - {}_1\theta_1). \tag{81}$$

Simultaneous solution of these equations yields improved estimates (θ_1', θ_2')

$$\theta_1' = \frac{{}_1\theta_1({}_2\theta_1 - {}_1\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)} - \frac{{}_2\theta_1({}_1\theta_1 - {}_0\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)}, \quad (82)$$

$$\theta_2' = \frac{{}_1\theta_2({}_2\theta_1 - {}_1\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)} - \frac{{}_2\theta_2({}_1\theta_1 - {}_0\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)}. \quad (83)$$

This is equivalent to solving

$$\frac{\theta_1' - {}_1\theta_1}{{}_2\theta_1 - {}_1\theta_1} = \frac{0 - ({}_1\theta_1 - {}_0\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)}, \quad (84)$$

$$\frac{\theta_2 - {}_1\theta_2}{{}_2\theta_2 - {}_1\theta_2} = \frac{0 - ({}_1\theta_1 - {}_0\theta_1)}{({}_2\theta_1 - {}_1\theta_1)-({}_1\theta_1 - {}_0\theta_1)}, \quad (85)$$

or interpolating in the table (Table 1).

Table 1. Interpolation for improved estimates (θ_1', θ_2')

θ_2	θ_1 from $L_1(\theta_1,\theta_2)=0$	θ_1 from $L_2(\theta_1,\theta_2)=0$	Difference
${}_1\theta_2$	${}_1\theta_1$	${}_0\theta_1$	$({}_1\theta_1 - {}_0\theta_1)$
${}_2\theta_2$	${}_2\theta_1$	${}_1\theta_1$	$({}_2\theta_1 - {}_1\theta_1)$
θ_2'	θ_1'		0

We may take the improved estimates (θ_1', θ_2') as the estimates $(\hat{\theta}_1, \hat{\theta}_2)$, or repeat the whole procedure with (θ_1', θ_2') replacing the starting values $({}_0\theta_1, {}_0\theta_2)$. This is Method 1 for two parameter distributions.

The maximum likelihood estimates $(\hat{\theta}_1, \hat{\theta}_2)$ of the parameters (θ_1, θ_2) are the values satisfying simultaneously the two estimating equations. The asymptotic variance-covariance matrix of $(\hat{\theta}_1, \hat{\theta}_2)$ can be approximated by

$$\begin{bmatrix} \hat{V}(\hat{\theta}_1) & \hat{\mathrm{Cov}}(\hat{\theta}_1, \hat{\theta}_2) \\ \hat{\mathrm{Cov}}(\hat{\theta}_2, \hat{\theta}_1) & \hat{V}(\hat{\theta}_2) \end{bmatrix} = - \begin{bmatrix} \partial^2 L/\partial\theta_1^2 & \partial^2 L/\partial\theta_1\partial\theta_2 \\ \partial^2 L/\partial\theta_2\partial\theta_1 & \partial^2 L/\partial\theta_2^2 \end{bmatrix}^{-1}_{(\theta_1,\theta_2)=(\hat{\theta}_1,\hat{\theta}_2)} . \tag{86}$$

The second partial derivatives, and hence the estimated asymptotic variances and covariance, can be expressed in terms of $P(y;\ \theta)$, $p(y;\ \theta)$ in the same manner as the first derivatives $\partial L/\partial\theta_1$, $\partial L/\partial\theta_2$ are so expressed (cf. Equations 37, 40, 59; 77).

METHOD 2, HARTLEY'S METHOD

The second procedure to be considered for maximum likelihood estimation of distribution parameters from incomplete data will be designated as Method 2, Hartley's method. See Hartley (1958). For incomplete data of a very general type from a given distribution, Hartley's method involves the computation of "pseudo-frequencies" for variate values in the subsets of truncation and censorship, and solution of maximum likelihood estimating equations for complete data, in an iterative procedure using only tables of distribution areas and ordinates. The method applies to any distribution, discrete or continuous, whose maximum likelihood estimating equations for complete data are solvable.

Our discussion here will parallel that given above for Method 1, and reference should be made to that discussion for omitted details. Consider a parent distribution, unspecified as being discrete or continuous, with parameter vector $\theta = (\theta_1, \cdot \cdot \cdot, \theta_s, \cdot \cdot \cdot, \theta_S)$ as described in Equations 24a, 24b, 25. Incomplete data, general case, is defined above with the aid of Equations 26, 27, 28a, 28b. For simplicity we consider distributions with but a single subset of censorship. Our sample of size n consists of, say, m counts $(0 \leq m \leq n)$ from the subset of censorship C and $n' = n - m$ values from the untruncated, uncensored set $R' = R-T-C$. In both the discrete and continuous cases, the maximum likelihood equations for estimating the parameter vector $\theta = (\theta_1, \cdot\cdot\cdot, \theta_s, \cdot\cdot\cdot, \theta_S)$ from incomplete (truncated and censored) data are (cf. Equation 32)

$$0 = \frac{\partial L}{\partial \theta_s} = -n \frac{\partial}{\partial \theta_s} \ln[1 - \Pr(Y \epsilon T;\ \theta)] + m \frac{\partial}{\partial \theta_s} \ln \Pr(Y \epsilon C;\ \theta)$$
$$+ \sum_{i=1}^{n'} \frac{\partial}{\partial \theta_s} \ln p(x_i;\ \theta), \qquad s = 1, 2, \cdot\cdot\cdot, S. \tag{87}$$

Let f_y represent the observed frequency of the variate value y for $y \epsilon R' = R-T-C$. In the continuous case, by using a quadrature formula approximation, we can write

$$\Pr(Y \epsilon T;\ \theta) = \sum_{y \epsilon T} w_y\ p(y;\ \theta), \tag{88}$$

$$\Pr(Y \epsilon C;\ \theta) = \sum_{y \epsilon C} w_y\ p(y;\ \theta), \tag{89}$$

where the w_y are "weights". In the discrete case the above expressions are exact with all w_y equal to unity. When these expressions are introduced into the estimating equations, we have (cf. Equation 37)

$$0 = \frac{\partial L}{\partial \theta_s} = n[1 - \sum_{y \epsilon T} w_y\ p(y;\ \theta)]^{-1}[\sum_{y \epsilon T} w_y\ \partial p(y;\ \theta)/\partial \theta_s]$$
$$+ m[\sum_{y \epsilon C} w_y\ p(y;\ \theta)]^{-1}[\sum_{y \epsilon C} w_y\ \partial p(y;\ \theta)/\partial \theta_s] \tag{90}$$
$$+ \sum_{y \epsilon R'} f_y \frac{\partial}{\partial \theta_s} \ln p(y;\ \theta), \qquad s = 1, 2, \cdot\cdot\cdot, S.$$

Now define "pseudo-frequencies" f'_y for $y \epsilon T$, C by "proportional allocation" as follows

$$f'_y = n[1 - \sum_{z\epsilon T} w_z\, p(z;\, \theta)]^{-1}[w_y\, p(y;\, \theta)], \qquad y\epsilon T, \tag{91}$$

$$f'_y = m[\sum_{z\epsilon C} w_z\, p(z;\, \theta)]^{-1}[w_y\, p(y;\, \theta)], \qquad y\epsilon C. \tag{92}$$

Substitution of these quantities into the estimating equations gives

$$\begin{aligned} 0 = \frac{\partial L}{\partial \theta_s} = \sum_{y\epsilon T} f'_y \frac{\partial}{\partial \theta_s} \ln p(y;\, \theta) + \sum_{y\epsilon C} f'_y \frac{\partial}{\partial \theta_s} \ln p(y;\, \theta) \\ + \sum_{y\epsilon R'} f_y \frac{\partial}{\partial \theta_s} \ln p(y;\, \theta), \quad s = 1, 2, \cdots, S. \end{aligned} \tag{93}$$

This is the form of the maximum likelihood equations for complete data. Hartley's iterative maximum likelihood estimating procedure is, then, as follows.

1. From an initial estimate ${}_0\theta = ({}_0\theta_1, \cdot\cdot\cdot, {}_0\theta_s, \cdot\cdot\cdot, {}_0\theta_S)$ of θ, find the pseudo-frequencies $f'_y = f'_y({}_0\theta)$ for $y\epsilon T$, C.

2. Using the observed frequencies f_y, $y\epsilon R' = R-T-C$, and the pseudo-frequencies f'_y for $y\epsilon T$, C from Step 1, solve the estimating equations for completed data, $0 = \partial L/\partial\theta_s = \partial L(\theta;\, {}_0\theta)/\partial\theta_s$, $s = 1, 2, \cdot\cdot\cdot, S$, for an improved estimate ${}_1\theta = ({}_1\theta_1, \cdot\cdot\cdot, {}_1\theta_s, \cdot\cdot\cdot, {}_1\theta_S)$ of θ. Continue Steps 1 and 2 until convergence ${}_0\theta, {}_1\theta, \cdot\cdot\cdot \longrightarrow \hat{\theta}$ such that $\partial L(\hat{\theta})/\partial\theta_s = \partial L(\theta;\, \hat{\theta})/\partial\theta_s \Big|_{\theta=\hat{\theta}} = 0$, $s = 1, 2, \cdot\cdot\cdot, S$.

This is Method 2, Hartley's method.

Hartley (1958) describes a method for obtaining estimates of the variances and covariances of the maximum likelihood estimates $\hat{\theta} = (\hat{\theta}_1, \cdot \cdot \cdot, \hat{\theta}_s, \cdot \cdot \cdot, \hat{\theta}_S)$. We consider only the cases of one and two coordinate parameter vectors θ. In the one parameter (one coordinate) case, if ${}_h\theta$ is an iteration estimate of the scalar θ, but not equal to $\hat{\theta}$, then $dL({}_h\theta)/d\theta = dL(\theta;\ {}_h\theta)/d\theta\Big|_{\theta={}_h\theta} \neq 0$, and $d^2L/d\theta^2$ can be approximated at the point $\theta = \hat{\theta}$ by

$$[d^2\tilde{L}/d\theta^2]_{\theta=\hat{\theta}} = [dL(\hat{\theta})/d\theta - dL({}_h\theta)/d\theta]/(\hat{\theta} - {}_h\theta) \tag{94}$$

$$= -[dL({}_h\theta)/d\theta]/(\hat{\theta} - {}_h\theta),$$

whence the variance of the maximum likelihood estimate $\hat{\theta}$ can be approximated by (cf. Equation 76)

$$\tilde{V}(\hat{\theta}) = (\hat{\theta} - {}_h\theta)/[dL({}_h\theta)/d\theta]. \tag{95}$$

In the two parameter case, $\theta = (\theta_1, \theta_2)$, the asymptotic variance-covariance matrix of $(\hat{\theta}_1, \hat{\theta}_2)$ can be estimated by (cf. Equation 86)

$$-\begin{bmatrix} \partial^2L/\partial\theta_1^2 & \partial^2L/\partial\theta_1\partial\theta_2 \\ \partial^2L/\partial\theta_2\partial\theta_1 & \partial^2L/\partial\theta_2^2 \end{bmatrix}^{-1}_{\theta=\hat{\theta}}$$

$$= -\begin{bmatrix} \partial\theta_1/\partial L_1 & \partial\theta_1/\partial L_2 \\ \partial\theta_2/\partial L_1 & \partial\theta_2/\partial L_2 \end{bmatrix}_{\theta=\hat{\theta}}, \tag{96}$$

where $L_1 = \partial L/\partial\theta_1$ and $L_2 = \partial L/\partial\theta_2$. Finite differences are used to obtain approximations $\partial\tilde{\theta_1}/\partial L_1$, $\partial\tilde{\theta_1}/\partial L_2 = \partial\tilde{\theta_2}/\partial L_1$,

$\partial\tilde{\theta_2}/\partial L_2$ of $\partial\theta_1/\partial L_1$, $\partial\theta_1/\partial L_2 = \partial\theta_2/\partial L_1$, $\partial\theta_2/\partial L_2$ at the point $(\theta_1, \theta_2) = (\hat{\theta}_1, \hat{\theta}_2)$, whence we have the estimated variance-covariance matrix of $(\hat{\theta}_1, \hat{\theta}_2)$ as

$$\begin{bmatrix} \tilde{V}(\hat{\theta}_1) & C\tilde{o}v(\hat{\theta}_1, \hat{\theta}_2) \\ C\tilde{o}v(\hat{\theta}_2, \hat{\theta}_1) & \tilde{V}(\hat{\theta}_2) \end{bmatrix} = - \begin{bmatrix} \partial\tilde{\theta_1}/\partial L_1 & \partial\tilde{\theta_1}/\partial L_2 \\ \partial\tilde{\theta_2}/\partial L_1 & \partial\tilde{\theta_2}/\partial L_2 \end{bmatrix} . \tag{97}$$

The method is described in detail in Hartley (1958).

APPLICATION OF METHODS 1, 2

Application to the Poisson Distribution

The Poisson parent distribution is defined in Equation 41. For incomplete (truncated and censored) data with a single subset of censorship, we have $T \subsetneqq R = \{0, 1, 2, \cdot \cdot \cdot\}$ and $C \subsetneqq R-T$, T and C otherwise arbitrary. A sample of n observations consists of, say, m counts from $C(0 \leq m \leq n)$ and $n' = n-m$ values from $R' = R-T-C$. Using the values of the Poisson derivatives $dp(y;\ \lambda)/d\lambda$, $d \ln p(y;\ \lambda)/d\lambda$ from Equation 42 in the general expression for the maximum likelihood estimating equation (Equation 37), we obtain the estimating equation for the Poisson parameter

$$0 = \frac{dL}{d\lambda} = n \frac{[-1 + \lambda^{-1} \sum_{y \epsilon T} yp(y;\ \lambda)]}{[1 - \sum_{y \epsilon T} p(y;\ \lambda)]}$$

$$+ m \frac{[\lambda^{-1} \sum_{y \epsilon C} yp(y;\ \lambda)]}{[\sum_{y \epsilon C} p(y;\ \lambda)]} + (n-m) \frac{\overline{x}}{\lambda}, \tag{98}$$

where $\overline{x} = (1/n') \sum_{i=1}^{n'} x_i$. For the special case where T and C consist of consecutive values, $T = \{t', t'+1, t'+2, \cdot \cdot \cdot, t''\}$ and $C = \{c', c'+1, c'+2, \cdot \cdot \cdot, c''\}$, we use the value for the Poisson derivative $dP(y;\ \lambda)/d\lambda$ from Equation 43 in the general expression, Equation 40, to obtain the simplified estimating equation

$$0 = \frac{dL}{d\lambda} = n \frac{[p(t'-1;\ \lambda) - p(t";\ \lambda)]}{[1 - P(t";\ \lambda) + P(t'-1;\ \lambda)]}$$

$$+ m \frac{[p(c'-1;\ \lambda) - p(c";\ \lambda)]}{[P(c";\ \lambda) - P(c'-1;\ \lambda)]} + (n-m)(\lambda^{-1}\bar{x}-1)\,. \tag{99}$$

The maximum likelihood estimate $\hat{\lambda}$ of the parameter λ is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{\lambda}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

Cohen (1954) gives the maximum likelihood estimating equations and asymptotic variance estimates for the special cases of incomplete data (1) singly truncated in the lower tail, $T = \{0, 1, 2, \cdots, t\}$, C = null; (2) singly truncated in the upper tail, $T = \{t, t+1, t+2, \cdots\}$, C = null; (3) doubly truncated in the tails, $T = \{0, 1, 2, \cdots, t'\} + \{t", t"+1, t"+2, \cdots\}$, C = null; (4) singly censored in the lower tail, T = null, $C = \{0, 1, 2, \cdots, c\}$; (5) singly censored in the upper tail, T = null, $C = \{c, c+1, c+2, \cdots\}$; (6) double censored in the tails combined, T = null, $C = \{0, 1, 2, \cdots, c'\} + \{c", c"+1, c"+2, \cdots\}$; (7) doubly censored in the tails separately, T = null, $C_1 = \{0, 1, 2, \cdots, c'\}$, $C_2 = \{c", c"+1, c"+2, \cdots\}$, m_1 counts from C_1, m_2 counts from C_2.

Hartley's method involves the computation of pseudo-frequencies f'_y for $y\epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y;\ \lambda)$ is obtained from Equation 41), and the solution of the maximum likelihood estimating equation (Equation 93), with the value of $d \ln p(y;\ \lambda)/d\lambda$ obtained from Equation 42. The estimating equation simplifies to

$$\lambda = \frac{\sum\limits_{y\epsilon T} yf'_y + \sum\limits_{y\epsilon C} yf'_y + \sum\limits_{y\epsilon R'} yf_y}{\sum\limits_{y\epsilon T} f'_y + \sum\limits_{y\epsilon C} f'_y + \sum\limits_{y\epsilon R'} f_y}\ . \tag{100}$$

Let ${}_h\lambda$ be an iteration estimate of λ. If ${}_h\lambda$ is substituted for λ in the expressions for f'_y, $y\epsilon T$, C, and these values are substituted in the right-hand side of Equation 100 to define the function $\emptyset({}_h\lambda)$, then Hartley's iterative procedure can be expressed compactly as

$${}_{h+1}\lambda = \emptyset({}_h\lambda)\,. \tag{101}$$

The maximum likelihood estimate $\hat{\lambda}$ is the limit approached by these iterates. Hartley's estimate of the variance of $\hat{\lambda}$ is obtained as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD".

Application to the Binomial Distribution with Numerical Example

The binomial parent distribution is defined in Equation 44. Let the incomplete data involve but a single subset of censorship. Using the values of the binomial derivatives $dp(y;\ N,p)/dp$, $d \ln p(y;\ N,p)/dp$ from Equation 45 in the general expression for the maximum likelihood estimating equation (Equation 37), we obtain the estimating equation for the binomial parameter p when N is known

$$0=\frac{(1-p)}{N}\frac{dL}{dp}=n\frac{[-1+(Np)^{-1}\sum_{y\epsilon T} yp(y;\ N,p)]}{[1-\sum_{y\epsilon T} p(y;\ N,p)]}$$

$$+\ m\frac{[(Np)^{-1}\sum_{y\epsilon C} yp(y;\ N,p)]}{[\sum_{y\epsilon C} p(y;\ N,p)]}+\frac{(n-m)\bar{x}}{Np}, \tag{102}$$

where $\bar{x} = (1/n')\sum_{i=1}^{n'} x_i$. For the special case where T and C consist of consecutive values, $T = \{t', t'+1, t'+2, \cdot\cdot\cdot, t''\}$ and $C = \{c', c'+1, c'+2, \cdot\cdot\cdot, c''\}$, we use the value of the binomial derivative $dP(y;\ N,p)/dp$ from Equation 46 in the general expression, Equation 40, to obtain the simplified estimating equation

$$0=\frac{1}{N}\frac{dL}{dp}=n\frac{[p(t'-1;\ N-1,p)-p(t'';\ N-1,p)]}{[1-P(t'';\ N,p)+P(t'-1;\ N,p)]}$$

$$+\ m\frac{[p(c'-1;\ N-1,p)-p(c'';\ N-1,p)]}{[P(c'';\ N,p)-P(c'-1;\ N,p)]}$$

$$+\ \frac{(n-m)}{(1-p)}[(Np)^{-1}\bar{x}-1]. \tag{103}$$

The maximum likelihood estimate $\hat{p}$ of the parameter p is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{\beta}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

Hartley's method involves the computation of pseudo-frequencies f'_y for $y\epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y;\ N,p)$ is obtained from Equation 44), and the solution of the maximum likelihood estimating equation (Equation 93), with the value of $d \ln p(y;\ N,p)/dp$ obtained from Equation 45. The estimating equation simplifies to

$$p = \frac{1}{N} \; \frac{\sum\limits_{y\epsilon T} yf'_y + \sum\limits_{y\epsilon C} yf'_y + \sum\limits_{y\epsilon R'} yf_y}{\sum\limits_{y\epsilon T} f'_y + \sum\limits_{y\epsilon C} f'_y + \sum\limits_{y\epsilon R'} f_y} . \tag{104}$$

The maximum likelihood estimate $\hat{p}$ of the binomial parameter p is the limit approached by the iterates ${}_hp$ in Hartley's two step iterative procedure. The asymptotic variance of $\hat{p}$ is estimated as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD".

Hartley (1958) gives the following example (Table 2) to illustrate the application of Method 2 to the complete censored binomial distribution.

Table 2. Binomial distribution, completely censored

	Variate values y			
	0 1 2	3 4 5	6 7 8	
f_y	- - -	- - -	- - -	
Observations	$m_1 = 14$	$m_2 = 73$	$m_3 = 19$	$n = 106$

It is desired to estimate the parameter p from the completely censored binomial distribution with parameters $N = 8$ and p. In the sample of $n = 106$ observations there are $m_1 = 14$ counts from $C_1 = \left\{c_1' = 0, 1, 2 = c_1''\right\}$, $m_2 = 73$ counts from $C_2 = \left\{c_2' = 3, 4, 5, = c_2''\right\}$, and $m_3 = 19$ counts from $C_3 = \left\{c_3' = 6, 7, 8 = c_3''\right\}$. $R = C_1 + C_2 + C_3$. Hartley's estimating equation is obtained from Equations 92, 44; 93, 45 and is

$$0 = (1-p)\frac{dL}{dp} = \frac{1}{p}\left[\sum_{y\epsilon C_1} yf_y' + \sum_{y\epsilon C_2} yf_y' + \sum_{y\epsilon C_3} yf_y'\right]$$
$$- N\left[\sum_{y\epsilon C_1} f_y' + \sum_{y\epsilon C_2} f_y' + \sum_{y\epsilon C_3} f_y'\right]. \tag{105}$$

This simplifies to (cf. Equation 104)

$$p = \frac{1}{N}\,\frac{\sum_{y\epsilon C_1} yf_y' + \sum_{y\epsilon C_2} yf_y' + \sum_{y\epsilon C_3} yf_y'}{\sum_{y\epsilon C_1} f_y' + \sum_{y\epsilon C_2} f_y' + \sum_{y\epsilon C_3} f_y'}\,. \tag{106}$$

The results of Hartley's iterative procedure are summarized in Table 3. We obtain $\hat{p} = 0.518$. The variance of $\hat{p}$ is estimated from Equation 95. In that expression ${}_hp = {}_1p = 0.5160$ and $\hat{p} = 0.5177$. The derivative $dL({}_hp)/dp$ is obtained from Equations 92, 105. We find

$$\tilde{V}(\hat{p}) = 0.000353. \tag{107}$$

The same sample will be worked by Method 1. From Equation 103 we write the estimating equation as

$$0 = \frac{1}{N}\frac{dL}{dp} = -m_1\,\frac{p(c_1''; N-1,p)}{P(c_1''; N,p)} + m_2\,\frac{[p(c_2'-1; N-1,p) - p(c_2''; N-1,p)]}{[P(c_2''; N,p) - P(c_2'-1; N,p)]}$$
$$+ m_3\,\frac{p(c_3'-1; N-1,p)}{[1 - P(c_3'-1; N,p)]}\,. \tag{108}$$

Table 3. Method 2, Hartley's method, to obtain $\hat{p}$

Iterate h	Estimate ${}_hp$	Pseudo-frequencies ${}_hf'_y$ $y=0$	1	2	3	4	5	6	7	8	Total
0		1	4	9	23	26	24	12	5	2	106
1	0.516										
1		0.34	2.89	10.77	21.04	28.04	23.92	14.03	4.30	0.57	105.90
2	0.5174										
2		0.34	2.88	10.78	20.92	28.03	24.05	14.10	4.33	0.57	106.00
3	0.5177										
	$\hat{p} = 0.518$										

The results of Method 1 are summarized in Table 4. We obtain by interpolation $\hat{p} = 0.518$.

Table 4. Method 1 to obtain $\hat{p}$

Trial p	dL(p)/dp
.5	45.405 421
.52	- 5.403 568
$\hat{p}$ = .5179	0
.51	20.010 874

The variance of $\hat{p}$ is estimated from Equation 76. In that expression d^2L/dp^2 is obtained by differentiating Equation 40. With the help of Equations 45, 46, d^2L/dp^2 can be written as

$$\frac{d^2L}{dp^2} = m_1\ N(N-1)\ \frac{[-p(c_1''-1;\ N-2,p) + p(c_1'';\ N-2,p)]}{P(c_1'';\ N,p)}$$

$$- m_1\ N^2\ \frac{[p(c_1'';\ N-1,p)]^2}{[P(c_1'';\ N,p)]^2}$$

$$+ m_2\ N(N-1)\ \frac{[p(c_2'-2;\ N-2,p) - p(c_2'-1;\ N-2,p) - p(c_2''-1;\ N-2,p) + p(c_2'';\ N-2,p)]}{[P(c_2'';\ N,p) - P(c_2'-1;\ N,p)]}$$

$$- m_2\ N^2\ \frac{[p(c_2'-1;\ N-1,p) - p(c_2'';\ N-1,p)]^2}{[P(c_2'';\ N,p) - P(c_2'-1;\ N,p)]^2}$$

$$+ m_3 \; N(N-1) \; \frac{[p(c_3'-2;\; N-2,p) \; - \; p(c_3'-1;\; N-2,p)]}{[1-P(c_3'-1;\; N,p)]}$$

$$- m_3 \; N^2 \; \frac{[p(c_3'-1;\; N-1,p)]^2}{[1 \; - \; P(c_3'-1;\; N,p)]^2} \; . \tag{109}$$

We find

$$\hat{V}(\hat{p}) = 0.000\;393. \tag{110}$$

Application to the Negative Binomial Distribution

The negative binomial parent distribution is defined in Equation 47. Let the incomplete data involve but a single subset of censorship. Without special aid tables of sums of reciprocals, the expressions for $\partial p(y;\; r,p)/\partial r$ and $\partial P(y;\; r,p)/\partial r$ given above (Equations 50, 51) are not simple, except in the case where $y = 0$. As a consequence, the maximum likelihood estimating equation $0 = \partial L/\partial r$ is not a simple expression unless $T = \{0\}$ and C = null. Therefore, estimation by Method 1 involving the parameter r will be discussed only for this case. Using the values of the negative binomial derivatives $dp(y;\; r,p)/dp$, $d \ln p(y;\; r,p)/dp$ from Equation 48 in the general expression for the maximum likelihood estimating equation (Equation 37), we obtain the estimating equation for the negative binomial parameter p when r is known

$$0 = -\frac{p}{r}\frac{dL}{dp} = n \; \frac{[-1 + r^{-1}(1-p)^{-1}p \sum\limits_{y \in T} y p(y;\; r,p)]}{[1 - \sum\limits_{y \in T} p(y;\; r,p)]}$$

$$+ m \frac{r^{-1}(1-p)^{-1}p \sum_{y \in C} yp(y;\ r,p)}{[\sum_{y \in C} p(y;\ r,p)]} + \frac{(n-m)p}{r(1-p)} \overline{x}, \tag{111}$$

where $\overline{x} = (1/n') \sum_{i=1}^{n'} x_i$. For the special case where T and C consist of consecutive values, $T = \{t', t'+1, t'+2, \cdot \cdot \cdot, t''\}$ and $C = \{c', c'+1, c'+2, \cdot \cdot \cdot, c''\}$, we use the value of the negative binomial derivative $dP(y;\ r,p)/dp$ from Equation 49 in the general expression, Equation 40, to obtain the simplified estimating equation

$$0 = \frac{-p}{r} \frac{dL}{dp} = n \frac{[P(t'-1;\ r,p) - P(t'';\ r,p) - P(t'-2;\ r+1,p) + P(t''-1;\ r+1,p)]}{[1 - P(t'';\ r,p) + P(t'-1;\ r,p)]}$$

$$+ m \frac{[P(c'-1;\ r,p) - P(c'';\ r,p) - P(c'-2;\ r+1,p) + P(c''-1;\ r+1,p)]}{[P(c'';\ r,p) - P(c'-1;\ r,p)]}$$

$$+ (n-m)[r^{-1}(1-p)^{-1}\ p\ \overline{x} - 1]. \tag{112}$$

The maximum likelihood estimate $\hat{p}$ of the parameter p is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{p}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

For the special case in which $T = \{0\}$ and C = null we consider the joint estimation by Method 1 of the two parameters p,r. The estimating equations for this case are (cf. Equations 78, 79)

$$0 = \frac{\partial L}{\partial p} = \frac{nr}{p(1-p^r)} - \frac{n\bar{x}}{(1-p)}, \tag{113}$$

$$0 = \frac{\partial L}{\partial r} = \frac{n \ln p}{(1-p^r)} + \sum_{i=1}^{n} \left[\sum_{z=0}^{x_i-1} 1/(r+z) \right], \tag{114}$$

where $(1-p^r) = 1 - p(0;\ r,p)$. The maximum likelihood estimates $\hat{p},\hat{r}$ of the parameters p,r are the values satisfying simultaneously the two estimating equations. The asymptotic variance-covariance matrix of $\hat{p},\hat{r}$ can be approximated by (cf. Equation 86)

$$\begin{bmatrix} \hat{V}(\hat{p}) & \hat{Cov}(\hat{p},\hat{r}) \\ \hat{Cov}(\hat{r},\hat{p}) & \hat{V}(\hat{r}) \end{bmatrix} = - \begin{bmatrix} \partial^2 L/\partial p^2 & \partial^2 L/\partial p \partial r \\ \partial^2 L/\partial r \partial p & \partial^2 L/\partial r^2 \end{bmatrix}^{-1}_{(p,r)=(\hat{p},\hat{r})}, \tag{115}$$

where

$$\frac{\partial^2 L}{\partial p^2} = \frac{nr[(r+1)p^r-1]}{p^2(1-p^r)^2} + \frac{n\bar{x}}{(1-p)^2}, \tag{116}$$

$$\frac{\partial^2 L}{\partial p \partial r} = \frac{n[rp^r \ln p - p^r + 1]}{p(1-p^r)^2}, \tag{117}$$

$$\frac{\partial^2 L}{\partial r^2} = \frac{np^r(\ln p)^2}{(1-p^r)^2} - \sum_{i=1}^{n} \left[\sum_{z=0}^{x_i-1} 1/(r+z)^2 \right]. \tag{118}$$

These estimating equations and the variance-covariance estimates were given by David and Johnson (1952) and also by Sampford (1955).

Hartley's method involves the computation of pseudo-frequencies f'_y for $y \epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y;\ r,p)$ is obtained from Equation 47), and the solution of the maximum likelihood estimating equations (Equations 93), with the values of $\partial \ln p(y;\ r,p)/\partial p$ and $\partial \ln p(y;\ r,p)/\partial r$ obtained from Equations 48, 50. The estimating equations simplify to

$$0 = -\frac{p}{r}\frac{\partial L}{\partial p} = \frac{p}{r(1-p)}\left[\sum_{y\epsilon T} y f'_y + \sum_{y\epsilon C} y f'_y + \sum_{y\epsilon R'} y f_y\right]$$

$$-\left[\sum_{y\epsilon T} f'_y + \sum_{y\epsilon C} f'_y + \sum_{y\epsilon R'} f_y\right], \qquad (119)$$

$$0 = \frac{\partial L}{\partial r} = (\ln p)\left[\sum_{y\epsilon T} f'_y + \sum_{y\epsilon C} f'_y + \sum_{y\epsilon R'} f_y\right]$$

$$+ \left[\sum_{y\epsilon T} f'_y \left(\sum_{z=0}^{y-1} 1/(r+z)\right) + \sum_{y\epsilon C} f'_y \left(\sum_{z=0}^{y-1} 1/(r+z)\right)\right.$$

$$\left. + \sum_{y\epsilon R'} f_y \left(\sum_{z=0}^{y-1} 1/(r+z)\right)\right]. \qquad (120)$$

The maximum likelihood estimates $\hat{p},\hat{r}$ of the parameters p,r are the limits approached by the iterates ${}_hp$, ${}_hr$ in Hartley's two step iterative procedure. The asymptotic variance-covariance matrix of $\hat{p},\hat{r}$ is estimated as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD". Hartley (1958) gives a worked example of two parameter estimation from the negative binomial distribution with zero-class missing. He compares his result with that of Sampford (1955) for the same example.

Application to the Geometric Distribution

The geometric parent distribution is defined in Equation 52. Let the incomplete data involve but a single subset of censorship. Using the values of the geometric derivatives $dp(y;\ \theta)/d\theta$, $d \ln p(y;\ \theta)/d\theta$ from Equation 53 in the general expression for the maximum likelihood estimating equation (Equation 37), we obtain the estimating equation for the geometric parameter θ

$$0 = -\theta(1-\theta)\frac{dL}{d\theta} = n\frac{[-1 + \theta \sum_{y\epsilon T} yp(y;\ \theta)]}{[1 - \sum_{y\epsilon T} p(y;\ \theta)]} + m\frac{[\theta \sum_{y\epsilon C} yp(y;\ \theta)]}{[\sum_{y\epsilon C} p(y;\ \theta)]} + (n-m)\theta\overline{x}, \qquad (121)$$

where $\overline{x} = (1/n') \sum_{i=1}^{n'} x_i$. For the special case where T and C consist of consecutive values, $T = \{t',\ t'+1,\ t'+2,\ \cdot\ \cdot\ \cdot,\ t''\}$ and $C = \{c',\ c'+1,\ c'+2,\ \cdot\ \cdot\ \cdot,\ c''\}$, we use the value of the geometric derivative $dP(y;\ \theta)/d\theta$ from Equation 54 in the general expression, Equation 40, to obtain the simplified estimating equation

$$0 = -\theta(1-\theta)\frac{dL}{d\theta} = n\theta\,\frac{[(t'-1)[1 - P(t'-1;\ \theta)]-t''[1 - P(t'';\ \theta)]}{[1 - P(t'';\ \theta) + P(t'-1;\ \theta)]}$$

$$+\ m\theta\,\frac{(c'-1)[1 - P(c'-1;\ \theta)] - c''[1 - P(c'';\ \theta)]}{[P(c'';\ \theta) - P(c'-1;\ \theta]}$$

$$+\ (n-m)(\theta\overline{x} - 1)\,. \tag{122}$$

The maximum likelihood estimate $\hat{\theta}$ of the parameter θ is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{\theta}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

Hartley's method involves the computation of pseudo-frequencies f'_y for $y \epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y;\ \theta)$ is obtained from Equation 52), and the solution of the maximum likelihood estimating equation (Equation 93), with the value of $d \ln p(y;\ \theta)/d\theta$ obtained from Equation 53. The estimating equation simplifies to

$$\theta = \frac{[\sum_{y\epsilon T} f'_y + \sum_{y\epsilon C} f'_y + \sum_{y\epsilon R'} f_y]}{[\sum_{y\epsilon T} yf'_y + \sum_{y\epsilon C} yf'_y + \sum_{y\epsilon R'} yf_y]}\ . \tag{123}$$

The maximum likelihood estimate $\hat{\theta}$ of the geometric distribution parameter θ is the limit approached by the iterates ${}_h\theta$ in Hartley's two step iterative procedure. The asymptotic variance of $\hat{\theta}$ is estimated as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD".

Application to the Normal Distribution with Numerical Example

The normal parent distribution is defined in Equation 60. Let the incomplete data involve single intervals of truncation and censorship. Using the values of the normal derivatives $\partial P(y;\ \mu,\sigma)/\partial\mu$, $\partial P(y;\ \mu,\sigma)/\partial\sigma$, $\partial \ln p(y;\ \mu,\sigma)/\partial\mu$, $\partial \ln p(y;\ \mu,\sigma)/\partial\sigma$ from Equations 62, 64, 61, 63 in the general expression for the maximum likelihood estimating equation (Equation 59), we obtain the estimating equations for the normal parameters μ, σ

$$0 = \sigma \frac{\partial L}{\partial \mu} = n\frac{[p(t'^*) - p(t''^*)]}{[1 - P(t''^*) + P(t'^*)]} + m\frac{[p(c'^*) - p(c''^*)]}{[P(c''^*) - P(c'^*)]} + (n-m)\ \frac{(\bar{x} - \mu)}{\sigma}, \tag{124}$$

$$0 = \sigma \frac{\partial L}{\partial \sigma} = n\ \frac{[t'^*p(t'^*) - t''^*p(t''^*)]}{[1 - P(t''^*) + P(t'^*)]} + m\ \frac{[c'^*p(c'^*) - c''^*p(c''^*)]}{[P(c''^*) - P(c'^*)]} + \sum_{i=1}^{n-m} [(x_i - \mu)/\sigma]^2 - (n-m), \tag{125}$$

where

$$y^* = (y - \mu)/\sigma \tag{126}$$

and $p(y) = p(y;\ 0,1)$, $P(y) = P(y;\ 0,1)$. The maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$ of the parameters μ, σ are the values satisfying simultaneously the two estimating equations. The asymptotic variance-covariance matrix of $\hat{\mu}$, $\hat{\sigma}$ is estimated as described in the section "For Two Parameter Distributions" of the chapter entitled "METHOD 1".

For the special case in which there is but a single finite point of truncation or censorship, and the two estimating equations are expressed in terms of the standard deviation σ and the normalized point, say, of truncation t''^* (t'^* being $-\infty$), it is possible to obtain a single equation in the one variable t''^* which is easily solved, and whose solution leads to estimates of μ and σ. See Cohen (1950).

Hartley's method involves the computation of pseudo-frequencies f_y' for $y \epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y;\ \mu,\ \sigma)$ is obtained from Equation 60), and the solution of the maximum likelihood estimating equations (Equations 93), with the values of $\partial \ln p(y;\ \mu,\sigma)/\partial\mu$ and $\partial \ln p(y;\ \mu,\sigma)/\partial\sigma$ obtained from Equations 61, 63. The proportional allocation equations (Equations 91, 92) can be given explicitly when the quadrature approximations of Equations 88, 89 are specified explicitly. (We require that the sets of truncation and censorship, T and C, be approximated by finite intervals.) The probabilities $\Pr(Y \epsilon T;\ \mu,\sigma)$ and $\Pr(Y \epsilon C;\ \mu,\sigma)$ of Equations 88, 89 are approximated by quadrature formulas. $Pr(Y\epsilon T;\ \mu,\sigma)$ is written

$$Pr(Y\epsilon T;\ \mu,\sigma) = \int_{t'}^{t''} p(y;\ \mu,\sigma)dy = (t''-t')\sum_{t=0}^{J} a_t p(y_t;\ \mu,\sigma), \tag{127}$$

where $T = (t',t'')$ and

$$y_0 = t';\quad y_t = y_0 + t(t''-t')/J,\quad t = 1, 2, \cdots, J, \tag{128}$$

and the coefficients a_t, such as those of Simpson's or Weddle's rule, satisfy

$$\sum_{t=0}^{J} a_t = 1. \tag{129}$$

$Pr(Y\ C;\ \mu,\sigma)$ is approximated similarly in terms weights $w_{y_c} = (c''-c')a_c$ applied to ordinates at $J + 1$ equally spaced points y_c in the interval of censorship C (cf. Equation 89). Equations 88, 89 become

$$f'_{y_t} = \frac{n(t''-t')a_t}{[1 - P(t''*) + P(t'*)]}\ p(y_t;\ \mu,\sigma),\ t = 0, 1, 2, \cdots, J, \tag{130}$$

$$f'_{y_c} = \frac{m(c''-c')a_c}{[P(c''*) - P(c'*)]}\ p(y_c;\ \mu,\sigma),\ c = 0, 1, 2, \cdots, J. \tag{131}$$

The estimating equations (Equations 93) simplify to

$$\mu = \frac{[\sum_{t=0}^{J} y_t f'_{y_t} + \sum_{c=0}^{J} y_c f'_{y_c} + \sum_{i=1}^{n-m} y_i f_{y_i}]}{[\sum_{t=0}^{J} f'_{y_t} + \sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}]} \tag{132}$$

$$\sigma^2 = \frac{\left[\sum_{t=0}^{J} (y_t-\mu)^2 f'_{y_t} + \sum_{c=0}^{J} (y_c-\mu)^2 f'_{y_c} + \sum_{i=1}^{n-m} (y_i-\mu) f_{y_i}\right]}{\left[\sum_{t=0}^{J} f'_{y_t} + \sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}\right]}, \tag{133}$$

where $f_{y_i} = 1$ for individual observed values y_i in R', $i = 1, 2, \cdots, (n-m)$.

Let $({}_h\mu, {}_h\sigma)$ be an iteration estimate of (μ, σ). If $({}_h\mu, {}_h\sigma)$ is substituted for (μ, σ) in the expressions for f'_y, $y \epsilon T$, C, and these values are substituted in the right-hand sides of Equations 132, 133 to define, respectively, the functions $\phi({}_h\mu, {}_h\sigma)$ and $\psi({}_h\mu, {}_h\sigma)$, then Hartley's iterative procedure can be expressed compactly as (cf. Equation 101)

$${}_{h+1}\mu = \phi({}_h\mu, {}_h\sigma), \tag{134}$$

$$({}_{h+1}\sigma^2) = \psi({}_h\mu, {}_h\sigma). \tag{135}$$

The maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$ are the limits approached by these iterates. The asymptotic variance-covariance matrix of $\hat{\mu}$, $\hat{\sigma}$ is estimated as described in the chapter entitled, "METHOD 2, HARTLEY'S METHOD".

Consider the following numerical example to illustrate the application of Methods 1 and 2 to the normal distribution. The example is adapted from data given by Kendall and Stuart (1958) on page 140. A sample of size n = 85 is drawn from a normal distribution with a single interval of censorship

$C = (64.5, 69.5)$. There occur $m = 58$ observations in C and $(n-m) = 27$ measured observations x_i, $i = 1, 2, \cdot \cdot \cdot, (n-m)$. These are: 61.0; 61.9, 62.1; 62.8, 62.9, 63.1, 63.2; 63.7, 63.8, 63.9, 64.0, 64.1, 64.2, 64.3; 69.7, 69.8, 69.9, 70.1, 70.2, 70.3; 70.8, 70.9, 71.1, 71.2; 71.9, 72.1; 73.0. For a convenient display of the data, we can round the values of observations x_i to whole numbers and obtain the grouped sample distribution as given in Table 5.

Table 5. Sample distribution (grouped for convenient display)

	Group mark													
	61	62	63	64	65	66	67	68	69	70	71	72	73	Total
Frequency	1	2	4	7	-	-	-	-	-	6	4	2	1	27
					m = 58									58
														n = 85

It is desired to estimate the parameters μ, σ of the censored normal distribution from the sample described above. The estimating equations for Method 1 are (cf. Equations 124, 125)

$$0 = \sigma \frac{\partial L}{\partial \mu} = m \frac{[p(c'^*) - p(c''^*)]}{[P(c''^*) - P(c'^*)]} + (n-m) \frac{(\bar{x}-\mu)}{\sigma}, \tag{136}$$

$$0 = \sigma\frac{\partial L}{\partial \sigma} = m \frac{[c'^*p(c'^*) - c''^*p(c''^*)]}{[P(c''^*) - P(c'^*)]}$$

$$+ \frac{1}{\sigma^2} [\sum_{i=1}^{n-m} x_i^2 - 2\mu \sum_{i=1}^{n-m} x_i + (n-m)\mu^2] -(n-m). \tag{137}$$

From the data, we have $c' = 64.5$, $c'' = 69.5$, $\sum_{i=1}^{n-m} x_i = 1806$, $\sum_{i=1}^{n-m} x_i^2 = 121,218.80$. The results of Method 1 obtained from an initial estimate ${}_0\mu = 66$ are summarized in Table 6 (cf. Table 1 above in the section "For Two Parameter Distributions" of the chapter entitled "METHOD 1"). We obtain $\hat{\mu} = 66.9557$, $\hat{\sigma} = 2.4806$. The variance-covariance matrix of $\hat{\mu}$, $\hat{\sigma}$ is

Table 6. Method 1, interpolation for estimates $\hat{\mu}$, $\hat{\sigma}$

σ	μ from $L_1(\mu,\sigma)=0$	μ from $L_2(\mu,\sigma)=0$	Difference
${}_1\sigma = 2.6744$	${}_1\mu = 66.9571$	${}_0\mu = 66.0000$	0.9571
${}_2\sigma = 2.4803$	${}_2\mu = 66.9557$	${}_1\mu = 66.9571$	-0.0014
$\sigma = 2.4806$	$\mu = 66.9557$		0

estimated from Equation 86. In that expression the derivatives $\partial^2L/\partial\mu^2$, $\partial^2L/\partial\mu\partial\sigma = \partial^2L/\partial\sigma\partial\mu$, $\partial^2L/\partial\sigma^2$ are obtained by differentiating Equation 59. With the help of Equations 61, 62, 63, 64, the derivatives can be written as

$$\sigma^2 \frac{\partial^2 L}{\partial\mu^2} = m\frac{[c'^*p(c'^*) - c''^*p(c''^*)]}{[P(c''^*) - P(c'^*)]} - m\frac{[p(c'^*) - p(c''^*)]^2}{[P(c''^*) - P(c'^*)]^2} - (n-m), \tag{138}$$

$$\sigma^2 \frac{\partial^2 L}{\partial\mu\partial\sigma} = m\frac{\left([(c'^*)^2-1]p(c'^*) - [(c''^*)^2-1]p(c''^*)\right)}{[P(c''^*) - P(c'^*)]}$$

$$- m \frac{[p(c'*) - p(c''*)][c'*p(c'*) - c''*p(c''*)]}{[P(c''*) - P(c'*)]^2}$$

$$- 2(n-m)(\overline{x}-\mu)/\sigma \tag{139}$$

$$\sigma^2 \frac{\partial^2 L}{\partial\sigma^2} = m \frac{\left(c'*[(c'*)^2-2]p(c'*) - c''*[(c''*)^2-2]p(c''*)\right)}{[P(c''*) - P(c'*)\ ]}$$

$$- m \frac{[c'*p(c'*) - c''*p(c''*)]^2}{[P(c''*) - P(c'*)]^2}$$

$$- \frac{3}{\sigma^2} [\sum_{i=1}^{n-m} x_i^2 - 2\mu \sum_{i=1}^{n-m} x_i + (n-m)\mu^2] + (n-m). \tag{140}$$

We find

$$\begin{bmatrix} \hat{V}(\hat{\mu}) & \hat{Cov}(\hat{\mu},\ \hat{\sigma}) \\ \hat{Cov}(\hat{\mu},\ \hat{\sigma}) & \hat{V}(\hat{\sigma}) \end{bmatrix} = \begin{bmatrix} 0.090650 & 0.000282 \\ 0.000282 & 0.037251 \end{bmatrix}. \tag{141}$$

The estimating equations for Method 2, Hartley's method, are obtained from Equations 131, 60; 132, 133 and are

$$\mu = \frac{[\sum_{c=0}^{J} y_c\ f'_{y_c} + \sum_{i=1}^{n-m} y_i\ f_{y_i}]}{[\sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}]}, \tag{142}$$

$$\sigma^2 = \frac{[\sum_{c=0}^{J} (y_c-\mu)^2 f'_{y_c} + \sum_{i=1}^{n-m} (y_i\ -\mu)^2 f_{y_i}]}{[\sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}]}. \tag{143}$$

The results of Hartley's iterative procedure, using Weddle's rule in the quadrature approximation of $Pr(Y\epsilon C;\ \mu,\ \sigma)$ of Equation 89 (cf. Equations 127, 128, 129), are summarized in Table 7. We obtain $\hat{\mu} = 66.9555$, $\hat{\sigma} = 2.4804$.

Application to the Gamma Distribution

The gamma parent distribution is defined in Equation 65. Let the incomplete data involve single intervals of truncation and censorship. The expression for $\partial P(y;\ \alpha,\ \beta)/\partial\alpha$ given above (Equation 67) cannot be expressed simply in terms of gamma distribution areas and ordinates, and, as a consequence, the maximum likelihood estimating equation $0 = \partial L/\partial\alpha$ cannot be so expressed. Hence estimation by Method 1 will not be considered for the parameter α. Using the values of the gamma derivatives $dP(y;\ \alpha,\ \beta)/d\beta$, $d \ln p(y;\ \alpha,\ \beta)/d\beta$ from Equations 69, 68 in the general expression for the maximum likelihood estimating equation (Equation 59), we obtain the estimating equation for the gamma scale parameter β when the parameter α is known

$$0 = \beta\,\frac{dL}{d\beta} = n\,\frac{[t'p(t';\ \alpha,\ \beta) - t''p(t'';\ \alpha,\ \beta)]}{[1 - P(t'';\ \alpha,\ \beta) + P(t';\ \alpha,\ \beta)]}$$

$$+ m\,\frac{[c'p(c';\ \alpha,\ \beta) - c''p(c'';\ \alpha,\ \beta)]}{[P(c'';\ \alpha,\ \beta) - P(c';\ \alpha,\ \beta)]}$$

$$+ (n-m)\bar{x}/\beta - (n-m)\alpha, \tag{144}$$

Table 7. Method 2, iteration for estimates $\hat{\mu}$, $\hat{\sigma}$

Iterate	Estimates			Pseudo-frequencies $_hf'_{Y_c}$							
			$c =$	0	1	2	3	4	5	6	
			$a_c =$	1/20	5/20	1/20	6/20	1/20	5/20	1/20	
h	$_h\mu$	$_h\sigma$	$Y_c =$	64.500	65.333	66.167	67.000	67.833	68.667	69.500	Total
0	66	2.7									
0				2.994	16.942	3.487	19.571	2.774	10.724	1.508	58.000
1	66.7921	2.6050									
1				2.280	14.354	3.263	20.085	3.100	12.961	1.955	57.998
2	66.9262	2.4868									
2				2.110	13.830	3.241	20.365	3.177	13.290	1.988	58.001
3	66.9499	2.4806									
3				2.086	13.738	3.232	20.382	3.189	13.370	2.003	58.000
4	66.9546	2.4804									
4				2.082	13.721	3.230	20.382	3.191	13.387	2.007	58.000
5	66.9555	2.4804									
	$\hat{\mu}$	$\hat{\sigma}$									

where $\bar{x} = (1/n') \sum_{i=1}^{n'} x_i$. The maximum likelihood estimate $\hat{\beta}$ of the parameter β is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{\beta}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

Des Raj (1953) gives the maximum likelihood estimating equations and the asymptotic variance estimates for the Pearson Type III distribution expressed in terms of the population mean μ, standard deviation σ, and third standard moment α_3. He considers incomplete data involving truncation in a single tail or in both tails, or censorship in a single tail or in both tails separately, or censorship in both tails combined. Most results require that α_3 be known. For the special case in which there is but a single finite point of truncation or censorship, it is possible to obtain a single equation in one variable which is easily solved, and whose solution leads to estimates of μ and σ, α_3 being known. The whole analysis is similar to that given by Cohen (1950) for the normal distribution.

den Broeder (1955) gives the maximum likelihood equation for the gamma distribution scale parameter β when α is known in the special cases of truncation in a single tail or censorship in a single tail.

Hartley's method involves the computation of pseudo-frequencies f'_y for $y \epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y; \alpha, \beta)$ is obtained from Equation 65), and the solution of the maximum likelihood estimating equations (Equations 93), with the values of $\partial \ln p(y; \alpha, \beta)/\partial\alpha$ and $\partial \ln p(y; \alpha, \beta)/\partial\beta$ obtained from Equations 66, 68. The estimating equations simplify to

$$0 = \frac{\partial L}{\partial \alpha} = [\sum_{t=0}^{J} f'_{y_t} \ln y_t + \sum_{c=0}^{J} f'_{y_c} \ln y_c + \sum_{i=1}^{n-m} f_{y_i} \ln y_i]$$

$$- [\sum_{t=0}^{J} f'_{y_t} + \sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}]\left(\ln \beta + \Gamma'(\alpha)/\Gamma(\alpha)\right), \tag{145}$$

$$0 = \frac{\beta}{\alpha}\frac{\partial L}{\partial \beta} = \frac{1}{\alpha\beta} [\sum_{t=0}^{J} y_t f'_{y_t} + \sum_{c=0}^{J} y_c f'_{y_c} + \sum_{i=1}^{n-m} y_i f_{y_i}]$$

$$- [\sum_{t=0}^{J} f'_{y_t} + \sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}], \tag{146}$$

where $f_{y_i} = 1$ for individual observed values y_i in R', $i = 1, 2, \cdot \cdot \cdot, (n-m)$. The maximum likelihood estimates $\hat{\alpha}$, $\hat{\beta}$ of the gamma distribution parameters α, β are the limits approached by the iterates ${}_h\alpha$, ${}_h\beta$ in Hartley's two step iterative procedure. The asymptotic variance-covariance matrix of $\hat{\alpha}$, $\hat{\beta}$ is estimated as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD".

Application to the Exponential Distribution

The exponential parent distribution is defined in Equation 70, and is a special case of the gamma distribution

(Equation 65). Let the incomplete data involve single intervals of truncation and censorship. Using the values of the exponential derivatives $dP(y;\ \theta)/d\theta$, $d \ln p(y;\ \theta/d\theta$ from Equations 72, 71 in the general expression for the maximum likelihood estimating equation (Equation 59), we obtain the estimating equation for the exponential parameter θ (cf. Equation 144 with $\alpha = 1$)

$$0 = \theta \frac{dL}{d\theta} = n \frac{[t'p(t';\ \theta) - t''p(t'';\ \theta)]}{[1 - P(t'';\ \theta) + P(t';\ \theta)]} + m \frac{[c'p(c';\ \theta) - c''p(c'';\ \theta)]}{[P(c'';\ \theta) - P(c';\ \theta)]} + (n-m)\overline{x}/\theta - (n-m), \tag{147}$$

where $\overline{x} = (1/n') \sum_{i=1}^{n'} x_i$. The maximum likelihood estimate $\hat{\theta}$ of the parameter θ is obtained as the solution of the estimating equation. An estimate of the variance of $\hat{\theta}$ is obtained as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

For the special case of truncation in the lower tail, or censorship in the upper tail, or both of these, it is possible to obtain an explicit solution for the estimating equation. Let $T = (0, t'')$ and $C = (c', \infty)$, so that we have truncation in the lower tail and censorship in the upper tail. Then $t' = 0$ and $P(t';\ \theta) = 0$, $t'p(t';\ \theta) = 0$; $c'' = \infty$ and $P(c'';\ \theta) = 1$, $c''p(c'';\ \theta) = 0$. The estimating equation becomes (cf.

Equation 147)

$$0 = \theta \frac{dL}{d\theta} = -n \frac{t''p(t''; \theta)}{[1 - P(t''; \theta)]} + m \frac{c'p(c'; \theta)}{[1 - P(c'; \theta)]}$$

$$+ (n-m)\bar{x}/\theta - (n-m). \tag{148}$$

Using the identity

$$[1 - P(y; \theta)] = \theta\, p(y; \theta), \tag{149}$$

we can solve the estimating equation explicitly for $\theta = \hat{\theta}$,

$$\hat{\theta} = \bar{x} - nt''/(n-m) + mc'/(n-m). \tag{150}$$

Deemer and Votaw (1955) give the maximum likelihood estimating equation and the asymptotic variance estimate for the exponential distribution scale parameter θ in the special cases of censorship or truncation in the upper tail.

Hartley's method involves the computation of pseudo-frequencies f'_y for $y \epsilon T$, C by proportional allocation from Equations 91, 92 (where $p(y; \theta)$ is obtained from Equation 70), and solution of the maximum likelihood estimating equation (Equation 93), with the value of $d \ln p(y; \theta)/d\theta$ obtained from Equation 71. The estimating equation simplifies to

$$\theta = \frac{[\sum_{t=0}^{J} y_t f'_{y_t} + \sum_{c=0}^{J} y_c f'_{y_c} + \sum_{i=1}^{n-m} y_i f_{y_i}]}{[\sum_{t=0}^{J} f'_{y_t} + \sum_{c=0}^{J} f'_{y_c} + \sum_{i=1}^{n-m} f_{y_i}]}, \tag{151}$$

where $f_{y_i} = 1$ for individual observed values y_i in R', $i = 1, 2, \cdot \cdot \cdot, (n-m)$. The maximum likelihood estimate $\hat{\theta}$ of the exponential distribution parameter θ is the limit approached by the iterates ${}_h\theta$ in Hartley's two step iterative procedure. The asymptotic variance of $\hat{\theta}$ is estimated as described in the chapter entitled "METHOD 2, HARTLEY'S METHOD".

Application to the Uniform Distribution

The uniform parent distribution is defined in Equation 73. It is easy to verify that only in the case in which one subset of censorship includes the upper tail of the distribution (for example, $C = (c'; \theta)$ is the likelihood function maximized at the point $\hat{\theta}$ which is the solution of the estimating equation $dL/d\theta = 0$. In other cases, the maximum likelihood estimate $\hat{\theta}$ is the smallest value that can be assigned to θ in the light of the information in the sample (with complete data, for instance, $\hat{\theta} = x_{max.}$). Let the incomplete data involve single intervals of truncation and censorship. We have $T = (t', t")$ and $C = (c', \theta)$. Using the values for the uniform derivatives $dP(y; \theta)/d\theta$, $d \ln p(y; \theta)/d\theta$ from Equations 75, 74, and the relation

$$dP(\theta; \theta)/d\theta = d(1)/d\theta = 0 \tag{152}$$

in the general expression for the maximum likelihood estimating equation (Equation 59), we obtain the estimating equation for the uniform parameter θ

$$0 = \theta \frac{dL}{d\theta} = n \frac{[t'p(t'; \theta) - t''p(t''; \theta)]}{[1 - P(t''; \theta) + P(t'; \theta)]}$$

$$+ m \frac{c'p(c';\ \theta)}{[1 - P(c';\ \theta)]} - (n-m). \tag{153}$$

On simplifying, we obtain the explicit solution $\theta = \hat{\theta}$,

$$\hat{\theta} = nc'/(n-m) - m(t'' - t')/(n-m). \tag{154}$$

The asymptotic variance of $\hat{\theta}$ can be approximated as described in the section "Variance of the Estimate" of the chapter entitled "METHOD 1".

Hartley's method is not applicable to the uniform distribution because for this distribution the estimate from complete data is not obtained by setting the derivative of the log likelihood equal to zero,

$$dL/d\theta = 0. \tag{155}$$

Nor in the special case $T = (t', t'')$, $C = (c', \theta)$ where the maximum likelihood estimate of the parameter θ is obtained by setting $dL/d\theta = 0$ does Hartley's method reduce the estimation problem from incomplete data to that from complete data, for under the approximation (cf. Equations 87, 89, 127)

$$\Pr(Y\epsilon C;\ \theta) = \sum_{c=0}^{J} (c'' - c')a_c\ p(y_c;\ \theta) \tag{156}$$

we have

$$d\Pr(Y\epsilon C;\ \theta)/d\theta \neq \sum_{c=0}^{J} (c'' - c')a_c\ dp(y_c;\ \theta)/d\theta, \tag{157}$$

since c'' in this case is a function of θ, $c'' = \theta$.

CONVERGENCE THEOREM

Hartley's Method

Consider a parent distribution, unspecified as being discrete or continuous, with parameter vector $\theta = (\theta_1, \cdots, \theta_s, \cdots, \theta_S)$ as described in Equations 24a, 24b, 25. Incomplete data, general case, is defined above with the aid of Equations 26, 27, 28a, 28b. Our sample of size n consists of, say, m_q counts $(0 \leq m_q)$ from the subset of censorship C_q, $q = 1, 2, \cdots, Q$ $(\Sigma\, m_q \leq n)$, and $n' = n - \Sigma\, m_q$ values from the untruncated, uncensored set $R' = R-T-\Sigma\, C_q$. The log likelihood function for the incomplete data (cf. Equation 29) is $L(\theta) = L$,

$$L = \ln\left(\text{const.}[1 - \Pr(Y \epsilon T;\ \theta)]^{-n} \prod_{q=1}^{Q} [\Pr(Y \epsilon C_q;\ \theta)]^{m_q} \prod_{i=1}^{n'} p(x_i;\ \theta)\right) \quad . \tag{158}$$

In both the discrete and continuous cases, the maximum likelihood equations for estimating the parameter vector $\theta = (\theta_1, \cdots, \theta_s, \cdots, \theta_S)$ from incomplete (truncated and censored) data are (cf. Equation 32)

$$0 = \frac{\partial L}{\partial \theta_s} = -\, n \frac{\partial}{\partial \theta_s} \ln[1 - \Pr(Y \epsilon T;\ \theta)]$$

$$+ \sum_{q=1}^{Q} m_q \frac{\partial}{\partial \theta_s} \ln \Pr(Y \epsilon C_q;\ \theta) + \sum_{i=1}^{n'} \frac{\partial}{\partial \theta_s} \ln p(x_i;\ \theta),$$

$$s = 1, 2, \cdots, S. \tag{159}$$

Let f_y represent the observed frequency of the variate value y for $y \epsilon R' = R-T-\Sigma C_q$. In the continuous case, by using a quadrature formula approximation, we can write

$$\Pr(Y \epsilon T;\ \theta) = \sum_{y \epsilon T} w_y\, p(y;\ \theta), \tag{160}$$

$$\Pr(Y \epsilon C_q;\ \theta) = \sum_{y \epsilon C_q} w_y\, p(y;\ \theta),\ q = 1, 2, \cdots, Q, \tag{161}$$

where the w_y are "weights". In the discrete case the above expressions are exact with all w_y equal to unity. It will sometimes be convenient to write the subset of truncation as $T = C_0$ and to double index the elements y of C_q as $y = (q, r)$, $q = 0, 1, 2, \cdots, Q$, $r = 1, 2, \cdots, R_q$. In this notation Equations 160, 161 can be written as

$$\Pr(Y \epsilon C_q;\ \theta) = \sum_{n=1}^{R_q} w_{qr}\, p(q,r;\ \theta),\ q = 0, 1, 2, \cdots, Q. \tag{162}$$

When the expressions of Equations 160, 161 are introduced into the estimating equations, we have (cf. Equation 37)

$$0 = \frac{\partial L}{\partial \theta_s} = n[1 - \sum_{y \epsilon T} w_y\, p(y;\ \theta)]^{-1}[\sum_{y \epsilon T} w_y \frac{\partial}{\partial \theta_s} p(y;\ \theta)]$$

$$+ \sum_{q=1}^{Q} m_q[\sum_{y \epsilon C_q} w_y\, p(y;\ \theta)]^{-1}[\sum_{y \epsilon C_q} w_y \frac{\partial}{\partial \theta_s} p(y;\ \theta)]$$

$$+ \sum_{y \epsilon R'} f_y \frac{\partial}{\partial \theta_s} \ln p(y;\ \theta), \quad s = 1, 2, \cdots, S. \tag{163}$$

Now define "pseudo-frequencies" f_y' for $y \epsilon T$, C_q by "Proportional allocation" as follows

$$f_y' = n[1 - \sum_{z \epsilon T} w_z\, p(z;\ \theta)]^{-1}[w_y\, p(y;\ \theta)], \quad y \epsilon T, \tag{164}$$

$$f'_y = m_q[\sum_{z\epsilon C_q} w_z\ p(z;\ \theta)]^{-1}[w_y\ p(y;\ \theta)],\ y\epsilon C_q,\qquad q = 1,\ 2,\ \cdot\ \cdot\ \cdot,\ Q. \tag{165}$$

Substitution of these quantities into the estimating equations gives

$$0 = \frac{\partial L}{\partial \theta_s} = \sum_{y\epsilon T} f'_y \frac{\partial}{\partial \theta_s} \ln p(y;\ \theta) + \sum_{q=1}^{Q} \sum_{y\epsilon C_q} f'_y \frac{\partial}{\partial \theta_s} \ln p(y;\ \theta)$$

$$+ \sum_{y\epsilon R'} f_y \frac{\partial}{\partial \theta_r} \ln p(y;\ \theta),\ s = 1,\ 2,\ \cdot\ \cdot\ \cdot,\ S. \tag{166}$$

This is the form of the maximum likelihood equations for complete data. Hartley's iterative maximum likelihood estimating procedure is, then, as follows.

1. From an initial estimate ${}_0\theta = ({}_0\theta_1, \cdot\ \cdot\ \cdot, {}_0\theta_s, \cdot\ \cdot\ \cdot, {}_0\theta_S)$ of θ, find the pseudo-frequencies $f'_y = f'_y({}_0\theta)$ for $y\epsilon T$, C_q.
2. Using the observed frequencies f_y, $y\epsilon R' = R\text{-}T\text{-}\Sigma\ C_q$, and the pseudo-frequencies f'_y for $y\epsilon T$, C_q from Step 1, solve the estimating equations for completed data, $0 = \partial L/\partial\theta_s = \partial L(\theta;\ {}_0\theta)/\partial\theta_s$, $s = 1, 2, \cdot\ \cdot\ \cdot, S$, for an improved estimate ${}_1\theta = ({}_1\theta_1, \cdot\ \cdot\ \cdot, {}_1\theta_s, \cdot\ \cdot\ \cdot, {}_1\theta_S)$ of θ. Continue Steps 1 and 2 until convergence: ${}_0\theta, {}_1\theta, \cdot\ \cdot\ \cdot \longrightarrow \hat{\theta}$ such that $\partial L(\hat{\theta})/\partial\theta_s = 0$, $s = 1, 2, \cdot\ \cdot\ \cdot, S$.

This is Method 2, Hartley's method.

Convergence of the Iterative Procedure

Hypotheses

The following hypotheses A, B are sufficient to ensure the convergence of Hartley's iterative procedure. This will be stated formally as a theorem below.

A. The partial derivatives $\partial p(y;\ \theta)/\partial\theta_s$, $\partial^2 p(y;\ \theta)/\partial\theta_s\partial\theta_\sigma$ exist and are continuous with respect to θ for all s, $\sigma = 1, 2, \cdot\cdot\cdot, S$, all $y \epsilon R$, and all θ in the parameter space U.

B. In the parameter space U there exists a bounded convex set U', and a point ${}_0\theta$ in the interior of U', such that, with $L^o(f_y;\ \theta)$ representing the log likelihood function for complete data

$$L^o(f_y;\ \theta) = \ln \prod_{y \epsilon R} [p(y;\ \theta)],^{f_y} \tag{167}$$

we have

1. $L^o(f_y;\ {}_0\theta) > L^o(f_y;\ \theta')$ for all θ' in the closure of the complement of U' and all f_y,

2. the quadratic form $- \sum_{s=1} \sum_{\sigma=1} L^o_{s\sigma}(f_y;\ \theta) z_s z_\sigma$ is positive definite for all θ in U' and all f_y, where

$$L^o_{s\sigma}(f_y;\ \theta) = \partial^2 L^o(f_y;\ \theta)/\partial\theta_s\ \partial\theta_\sigma, \tag{168}$$

3. the quadratic form involving the log likelihood function for incomplete data $- \sum_{s=1} \sum_{\sigma=1} L_{s\sigma}(\theta)\ z_s z_\sigma$ is positive definite for all θ

in U', where

$$L_{s\sigma}(\theta) = \partial^2 L(\theta)/\partial\theta_s\, \partial\theta_\sigma. \tag{169}$$

Theorem

If the frequency or probability density function $p(y;\ \theta)$ of the random variable Y is such that Hypotheses A, B are satisfied, then, given the incomplete (truncated and censored) data in a sample of size n with, say, m_q counts in C_q, the approximations ${}_h\theta$ of Hartley's iterative procedure converge to an estimate $\hat{\theta}$ of θ which satisfies the maximum likelihood equations for incomplete data.

Proof of the convergence theorem

There are four steps to the proof of the convergence theorem. We designate these by a, b, c, and d, and give an outline of the entire proof before examining in detail the individual steps.

a. The iterative procedure is equivalent to successive maximizations of a certain pseudo-log likelihood function $L^*(f_y;\ \theta)$,

$$L^*(f_y;\ \theta) = \ln \prod_{y\in R} [w_y\ p(y;\ \theta)/f_y]^{f_y}. \tag{170}$$

Step 1 of the iterative procedure is equivalent to maximizing $L^*(f_y;\ {}_h\theta)$ with respect to the f_y under the side conditions

$$\sum_{y\in T} f_y = n[1 - \Pr(Y\in T;\ {}_h\theta)]^{-1}[\Pr(Y\in T;\ {}_h\theta)], \tag{171}$$

$$\sum_{y\in C_q} f_y = m_q, \qquad q = 1, 2, \cdots, Q. \tag{172}$$

The resulting values ${}_hf_y$ of the f_y, $y \epsilon T$, C_q, give the unique absolute maximum of $L^*(f_y;\ {}_h\theta)$. Step 2 of the iterative procedure is equivalent to maximizing $L^*({}_hf_y;\ \theta)$ with respect to θ. By Hypothesis B,2, the resulting value ${}_{h+1}\theta$ of θ gives the unique absolute maximum of $L^*({}_hf_y;\ \theta)$. The sequence $\{L^*({}_hf_y;\ {}_h\theta)\}$ is monotonic non-decreasing and bounded from above, hence convergent.

$$\lim_{h \to \infty} L^*({}_hf_y;\ {}_h\theta) = \tilde{L} \tag{173}$$

b. From the boundedness of U' and the continuity assumptions, there exists in the interior of U' a limit point θ^* of the sequence $\{{}_h\theta\}$, and a subsequence $\{{}_g\theta\}$ of $\{{}_h\theta\}$ such that

$$\lim_{g \to \infty} {}_g\theta = \theta^*. \tag{174}$$

If f_y^* is evaluated from θ^* in the proportional allocation equations, $y \epsilon T$, C_q, and for each $g = 1, 2, \cdots$, the frequency ${}_gf_y$ is evaluated from ${}_g\theta$ in the proportional allocation equations, then, by the continuity assumptions,

$$\lim_{g \to \infty} {}_gf_y = f_y^* \tag{175}$$

for all $y \epsilon T$, C_q, and

$$L^*(f_y^*;\ \theta^*) = \lim_{g \to \infty} L^*({}_gf_y;\ {}_g\theta) = \tilde{L}. \tag{176}$$

c. Let ${}_gf_y$ bear the index $h = h(g)$ in the original sequence $\{{}_hf_y\}$. Corresponding to each ${}_gf_y$ of the sequence $\{{}_gf_y\}$ there is a parameter vector estimate ${}_{h(g)+1}\theta$, rewritten for simplicity as ${}_g\theta'$, satisfying the maximum likelihood estimating equations (Equations 166), or, equivalently,

$$0 = \partial L^*({}_g f_y;\ {}_g\theta')/\partial\theta_s = \partial L^o({}_g f_y;\ {}_g\theta')/\partial\theta_s,$$

$$s = 1, 2, \cdots, S. \qquad (177)$$

There exists in the interior of U' a limit point θ' of the sequence $\{{}_g\theta'\}$ and a subsequence $\{{}_f\theta'\}$ of $\{{}_g\theta'\}$ such that

$$\lim_{f \to \infty} {}_f\theta' = \theta'. \qquad (178)$$

For each f, we have that

$$0 = L_s^*({}_f f_y;\ {}_f\theta'), \qquad s = 1, 2, \cdots, S. \qquad (179)$$

By the continuity assumptions,

$$L_s^*(f_y^*;\ \theta') = \lim_{f \to \infty} L_s^*({}_f f_y;\ {}_f\theta') = 0. \qquad (180)$$

Thus θ' is the unique value of θ maximizing $L^*(f_y^*;\ \theta)$ in Step 2 of the iterative procedure. But

$$L^*(f_y^*;\ \theta') = \lim_{f \to \infty} L^*({}_f f_y;\ {}_f\theta') = \tilde{L} = L^*(f_y^*;\ \theta^*). \qquad (181)$$

It follows that $\theta' = \theta^*$, and further that

$$0 = L_s^*(f_y^*;\ \theta^*) = L_s(\theta^*), \ s = 1, 2, \cdots, S, \qquad (182)$$

where $L_s(\theta)$ is defined in Equation 159.

d. From the result $L_s(\theta^*) = 0$, $s = 1, 2, \cdots, S$, for any limit point θ^* of the sequence $\{{}_h\theta\}$, and the positive definiteness of the quadratic form $-\sum_{s=1}^{S}\sum_{\sigma=1}^{S} L_{s\sigma}(\theta) z_s z_\sigma$ for all θ in U' (a consequence of Hypothesis B,3), it follows that the sequence $\{{}_h\theta\}$ has a unique limit point θ^* in the bounded closure of U' and hence is convergent to θ^*. The sequence

$\{{}_hf_y\}$ is convergent to f_y^*. Finally, the expression

$$0 = L_s(\theta^*), \qquad s = 1, 2, \cdots, S, \tag{183}$$

states that θ^* satisfies the maximum likelihood equations for incomplete data, $\theta^* = \hat{\theta}$.

a. We now show that Step 1 of the iterative procedure is equivalent to maximizing $L^*(f_y;\ {}_h\theta)$ with respect to the f_y (cf. Equation 170) under the side conditions of Equations 171, 172. For fixed ${}_h\theta$, the right-hand side of Equation 171 is a constant, which we shall denote by $m_0 = m_0({}_h\theta)$. Incorporation of the side conditions will be effected by elimination of the dependent variates. Denote T by C_0 and let f_{q1} be eliminated from each set of $f_{qr} = f_y$, $y \epsilon C_q$. Then the function $L^*(f_y;\ {}_h\theta)$ can be written as

$$\begin{aligned} L^*(f_{qr};\ {}_h\theta) = & \sum_{q=0}^{Q} \sum_{r=2}^{R_q} f_{qr} \ln[w_{qr} p(q,r;\ {}_h\theta] \\ & + \sum_{q=0}^{Q} (m_q - \sum_{r=2}^{R_q} f_{qr}) \ln[w_{q1} p(q,1;\ {}_h\theta)] \\ & - \sum_{q=0}^{Q} \sum_{r=2}^{R_q} f_{qr} \ln f_{qr} - \sum_{q=0}^{Q} (m_q - \sum_{r=2}^{R_q} f_{qr}) \ln(m_q - \sum_{r=2}^{R_q} f_{qr}) \\ & + \sum_{i=1}^{n'} \ln p(x_i;\ {}_h\theta). \end{aligned} \tag{184}$$

When the partial derivatives of L^* with respect to the f_{qr} ($q = 0, 1, 2, \cdots, Q$, $r = 2, 3, \cdots, R_q$) are set equal to zero, we obtain

$$0 = \frac{\partial L^*}{\partial f_{qr}} = \ln[w_{qr}p(q,r;\ {}_h\theta)] - \ln[w_{q1}p(q,1;\ {}_h\theta)] - \ln f_{qr}$$

$$+ \ln\,(m_q - \sum_{r=2}^{R_q} f_{qr}),\quad q = 0, 1, 2, \cdot\cdot\cdot, Q,$$
$$r = 2, 3, \cdot\cdot\cdot, R_q. \qquad (185)$$

Equations 185 simplify to

$$f_{qr} = [w_{q1}p(q,1;\ {}_h\theta)]^{-1}(m_q - \sum_{r=2}^{R_q} f_{qr})w_{qr}p(q,r;\ {}_h\theta),$$

$$q = 0, 1, 2, \cdot\cdot\cdot, Q,\ r = 2, 3, \cdot\cdot\cdot, R_q. \qquad (186)$$

The f_{qr} of Equations 186 satisfy the side conditions of Equations 171, 172, from which we obtain the identity

$$m_q = [w_{q1}p(q,1;\ {}_h\theta)]^{-1}(m_q - \sum_{r=2}^{R_q} f_{qr})\Pr(Y \in C_q;\ {}_h\theta). \qquad (187)$$

For each $q = 0, 1, 2, \cdot\cdot\cdot, Q$, Equation 186 is satisfied for f_{q1}, that is, for $r = 1$. Therefore, with Equation 187 substituted in Equation 186, recalling that $m_0 = m_0({}_h\theta)$ represents the right-hand side of Equation 169, we obtain Equations 164, 165, the proportional allocation equations of Step 1 of the iterative procedure.

Next we show that the values ${}_hf_{qr}$ of the f_{qr} in Step 1 of the iterative procedure (Equations 186, 164, 165) give the unique absolute maximum of $L^*(f_{qr};\ {}_h\theta)$. With ${}_h\theta$ fixed, expand $L^*(f_{qr};\ {}_h\theta) = L^*(f_{qr})$ for an arbitrary point f_{qr} $(q = 0, 1, 2, \cdot\cdot\cdot, Q,\ r = 1, 2, \cdot\cdot\cdot, R_q)$ in a second order Taylor series about the point ${}_hf_{qr}$. We obtain

$$L^*(f_{qr}) = L^*({}_hf_{qr}) + \sum_{q,r}[\partial L^*({}_hf_{qr})/\partial f_{qr}](f_{qr} - {}_hf_{qr})$$

$$+ \frac{1}{2!}\sum_{q,r}\sum_{q',r'}[\partial^2 L^*(\tilde{f}_{qr})/\partial f_{qr}\,\partial f_{q'r'}](f_{qr} - {}_hf_{qr})(f_{q'r'}$$

$$- {}_hf_{q'r'}), \tag{188}$$

where $\tilde{f}_{qr}$ is on the line segment joining f_{qr}, ${}_hf_{qr}$. But by Equations 185, the second term on the right-hand side of Equation 188 is zero. By differentiating Equation 185 with respect to $f_{q'r'}$, we can obtain the second partial derivatives of L* with respect to the f_{qr}, and show that the quadratic form

$$-\sum_{q,r}\sum_{q'r'}[\partial^2 L^*(\tilde{f}_{qr})/\partial f_{qr}\,\partial f_{q'r'}]z_{qr}z_{q'r'} \tag{189}$$

is positive definite for all $\tilde{f}_{qr}$($q = 0, 1, 2, \cdot\cdot\cdot, Q$, $r = 1, 2, \cdot\cdot\cdot, R_q$). It follows from this result and Equation 188 that

$$L^*({}_hf_{qr}) - L^*(f_{qr}) > 0, \tag{190}$$

unless $f_{qr} = {}_hf_{qr}$ for all $q = 0, 1, 2, \cdot\cdot\cdot, Q$, $r = 1, 2, \cdot\cdot\cdot, R_q$. That is, the values ${}_ff_{qr}$ of f_{qr} in Step 1 of the iterative procedure give the unique absolute maximum of $L^*(f_{qr};\ {}_h\theta)$.

We shall show that Step 2 of the iterative procedure is equivalent to maximizing $L^*({}_hf_{qr};\ \theta)$ with respect to θ, and that the value ${}_{h+1}\theta$ of θ gives the unique absolute maximum of $L^*({}_hf_{qr};\ \theta)$. Let ${}_h\theta$ be given, $h \geq 0$, and let ${}_hf_{qr}$

$(q = 0, 1, 2, \cdots, Q, r = 1, 2, \cdots, R_q)$ satisfy Equations 171, 172. Then the solution for θ of the maximum likelihood equations for complete data (cf. Equation 166) $0 = \partial L^o({}_hf_{qr};\ \theta)/\partial\theta_s$, $s = 1, 2, \cdots, S$, is equivalent to the solution for θ (cf. Equations 170, 166) of

$$0 = L_s^*({}_hf_{qr};\ \theta) = \sum_{q=0}^{Q} \sum_{r=1}^{R_q} {}_hf_{qr} \frac{\partial}{\partial\theta_s} \ln p(q,r;\ \theta)$$

$$+ \sum_{i=1}^{n'} \frac{\partial}{\partial\theta_s} \ln p(x_i;\ \theta),\ s = 1, 2, \cdots, S. \tag{191}$$

By Hypothesis A, $L^*({}_hf_{qr};\ \theta)$ is continuous with respect to θ for all θ in the closure of U', closed and bounded. And $L_s^*({}_hf_{qr};\ \theta) = \partial L^*({}_hf_{qr};\ \theta)/\partial\theta_s$ exists for all θ in U' closure. By Hypothesis B,1, there exists a point ${}_0\theta$ in the interior of U' such that $L^o(f_{qr};\ {}_0\theta) > L^o(f_{qr};\ \theta')$ for all θ' in the closure of the complement of U'. It follows that $L^*({}_hf_{qr};\ \theta)$ assumes an absolute maximum at some point ${}_{h+1}\theta$ in the interior of U', and that ${}_{h+1}\theta$ is among the solutions of Equations 191. By Hypothesis A, the second partial derivatives $L^*_{s\sigma} = \partial^2L^*/\partial\theta_s\,\partial\theta_\sigma$ exist and are continuous with respect to θ for all θ in U'. With ${}_hf_{qr}$ fixed, expand $L^*({}_hf_{qr};\ \theta) = L^*(\theta)$ for an arbitrary point θ in U' in a second order Taylor series about the point ${}_{h+1}\theta$ in U'. We obtain

$$L^*(\theta) = L^*({}_{h+1}\theta) + \sum_{s=1}^{S} L_s^*({}_{h+1}\theta)(\theta_s - {}_{h+1}\theta_s)$$

$$+ \frac{1}{2!} \sum_{s=1}^{S} \sum_{\sigma=1}^{S} L^*_{s\sigma}(\tilde{\theta})(\theta_s - {}_{h+1}\theta_s)(\theta_\sigma - {}_{h+1}\theta_\sigma), \tag{192}$$

where $\tilde{\theta}$ is on the line segment joining θ and ${}_{h+1}\theta$, hence in U' by the convexity. But ${}_{h+1}\theta$ satisfies Equations 191, so that the second term on the right-hand side of Equation 192 is zero. From the positive definiteness of the quadratic form of Hypothesis B,2, it follows that

$$L^*({}_{h+1}\theta) - L^*(\theta) > 0, \tag{193}$$

unless $\theta_s = {}_{h+1}\theta_s$, $s = 1, 2, \cdot \cdot \cdot, S$. That is, the value ${}_{h+1}\theta$ of θ gives the unique absolute maximum of $L^*({}_h f_{qr};\ \theta)$.

From the successive maximizations of the steps of the iterative procedure, we have

$$L^*({}_0 f_{qr};\ {}_0\theta) \leq L^*({}_0 f_{qr};\ {}_1\theta) \leq L^*({}_1 f_{qr};\ {}_1\theta) \leq L^*({}_1 f_{qr};\ {}_2\theta) \leq \cdot \cdot \cdot . \tag{194}$$

Thus the sequence $\{L^*({}_h f_{qr};\ {}_h\theta\}$ is monotonic non-decreasing. It is easily verified that each $L^*({}_h f_{qr};\ {}_h\theta)$ is bounded from above by zero. Hence the sequence is convergent. We can write for some real number $\tilde{L}$,

$$\lim_{h \to \infty} L^*({}_h f_{qr};\ {}_h\theta) = \tilde{L}. \tag{195}$$

b. It has been shown that ${}_h\theta$ is in the interior of U' for all ${}_h\theta$ in the sequence $\{{}_h\theta\}$. There exists a limit point θ^* in the closure of U', closed and bounded, and a subsequence $\{{}_g\theta\}$ of $\{{}_h\theta\}$ such that

$$\lim_{g \to \infty} {}_g\theta = \theta^*. \tag{196}$$

If f^*_{qr} is evaluated from θ^* in place of θ in the proportional allocation equations, Equations 164, 165, and for each $g = 1, 2, \cdot \cdot \cdot$, the frequency ${}_g f_{qr}$ is evaluated from ${}_g\theta$ in these equations, then, by the continuity assumptions of Hypothesis A,

$$\lim_{g \to \infty} {}_g f_{qr} = f^*_{qr}, \quad q = 0, 1, 2, \cdot \cdot \cdot, Q, \quad r = 1, 2, \cdot \cdot \cdot, R_q. \tag{197}$$

It also follows from Hypothesis A and the form of Equation 170 that $L^*(f_{qr}; \theta)$ is continuous jointly with respect to f_{qr} and θ, whence

$$\tilde{L} = \lim_{h \to \infty} L^*({}_h f_{qr}; {}_h\theta) = \lim_{g \to \infty} L^*({}_g f_{qr}; {}_g\theta) = L^*(f^*_{qr}; \theta^*). \tag{198}$$

From Equation 198 and Hypothesis B,1, it follows that θ^* is in the interior of U'.

c. We now proceed to show that $0 = L^*_s(f^*_{qr}; \theta^*)$ for all $s = 1, 2, \cdot \cdot \cdot, S$. Let ${}_g f_{qr}$ bear the index h(g) in the original sequence $\{{}_h f_{qr}\}$. Corresponding to each ${}_g f_{qr}$ of the sequence $\{{}_g f_{qr}\}$ there is a parameter vector estimate ${}_{h(g)+1}\theta$, rewritten for simplicity as ${}_g\theta'$, satisfying the maximum likelihood equations (Equations 166), or, equivalently, Equations 191.

$$0 = L^*_s({}_g f_{qr}; {}_g\theta'), \quad s = 1, 2, \cdot \cdot \cdot, S. \tag{199}$$

There exists in the interior of U' a limit point θ' of the sequence $\{{}_g\theta'\}$, and a subsequence $\{{}_f\theta'\}$ of $\{{}_g\theta'\}$ such that

$$\lim_{f \to \infty} {}_f\theta' = \theta'. \tag{200}$$

Each ${}_{f}\theta' = {}_{h(f)+1}\theta$ corresponds to an ${}_{f}f_{qr}$ through Equation 199. From Equation 197

$$\lim_{f \to \infty} {}_{f}f_{qr} = \lim_{g \to \infty} {}_{g}f_{qr} = f^*_{qr}, \quad q = 0, 1, 2, \cdots, Q, \quad r = 1, 2, \cdots, R_q. \tag{201}$$

By the continuity assumptions of Hypothesis A and the form of Equations 191, it follows that $L^*_s(f_{qr};\, \theta)$ is continuous jointly with respect to f_{qr} and θ, whence

$$L^*_s(f^*_{qr};\, \theta') = \lim_{f \to \infty} L^*_s({}_{f}f_{qr};\, {}_{f}\theta') = 0, \quad s = 1, 2, \cdots, S. \tag{202}$$

Thus θ' is the unique value of θ maximizing $L^*(f^*_{qr};\, \theta)$ in Step 2 of the iterative procedure. But from Equations 194, 195, 198,

$$L^*(f^*_{qr};\, \theta') = \lim_{f \to \infty} L^*({}_{f}f_{qr};\, {}_{f}\theta') = \tilde{L} = L^*(f^*_{qr};\, \theta^*). \tag{203}$$

It follows that $\theta' = \theta^*$, and hence from Equation 202 that

$$L^*_s(f^*_{qr};\, \theta^*) = 0, \qquad s = 1, 2, \cdots, S. \tag{204}$$

As observed earlier, the maximum likelihood estimating equations (Equations 166) and Equations 191 are equivalent. When these equations are expressed in terms of θ^* and f^*_{qr}, the latter evaluated from θ^* in the proportional allocation equations (Equations 164, 165), then they express the relationship of Equations 163, 159--that θ^* satisfies the maximum likelihood estimating equations from incomplete data. Thus Equations 204 imply

$$0 = \frac{\partial L}{\partial \theta_s}(\theta^*) = L_s(\theta^*), \quad s = 1, 2, \cdots, S. \tag{205}$$

d. Finally, we show that the sequence $\{{}_h\theta\}$ has but a single limit point θ^*, and is convergent to θ^*. Let θ^* and θ^{**} be two limit points of $\{{}_h\theta\}$. Define

$$\theta_s = \theta_s^* + u(\theta_s^{**} - \theta_s^*), \quad 0 \leq u \leq 1, \quad s = 1, 2, \cdots, S, \tag{206}$$

and

$$\phi(u) = \sum_{s=1}^{S} L_s(\theta)(\theta_s^{**} - \theta_s^*), \tag{207}$$

where $\theta = (\theta_1, \cdots, \theta_s, \cdots, \theta_S)$. The function $\phi(u)$ is continuous with respect to u for all $0 \leq u \leq 1$, and $\phi(0) = \phi(1) = 0$ by Equations 205. By the continuity assumptions of Hypothesis A, $\phi'(u)$ exists for all $0 < u < 1$,

$$\phi'(u) = d[\sum_{s=1}^{S} L_s(\theta)(\theta_s^{**} - \theta_s^*)]/du$$

$$= \sum_{s=1}^{S} (\theta_s^{**} - \theta_s^*)[\sum_{\sigma=1}^{S} (\partial L_s(\theta)/\partial\theta_\sigma)(d\theta_\sigma/du)]$$

$$= \sum_{s=1}^{S} \sum_{\sigma=1}^{S} L_{s\sigma}(\theta)(\theta_s^{**} - \theta_s^*)(\theta_\sigma^{**} - \theta_\sigma^*). \tag{208}$$

Therefore, by Rolle's theorem, there exists a real number $\tilde{u}$, $0 < \tilde{u} < 1$, such that

$$\phi'(\tilde{u}) = 0. \tag{209}$$

If we write $\tilde{\theta} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_s, \cdots, \tilde{\theta}_S)$, where

$$\tilde{\theta}_s = \theta_s^* + \tilde{u}(\theta_s^{**} - \theta_s^*), \quad s = 1, 2, \cdots, S, \tag{210}$$

then

$$0 = \phi'(\tilde{u}) = \sum_{s=1}^{S} \sum_{\sigma=1}^{S} L_{s\sigma}(\tilde{\theta})(\theta_s^{**} - \theta_s^*)(\theta_\sigma^{**} - \theta_\sigma^*). \tag{211}$$

But θ^* and θ^{**} are in the convex set U', so that $\tilde{\theta}$ is in U'. By Hypothesis B,3, the quadratic form in Equation 211 is negative definite. Therefore

$$(\theta_s^{**} - \theta_s^*) = 0, \quad s = 1, 2, \cdot \cdot \cdot, S. \tag{212}$$

That is, $\theta^{**} = \theta^*$ and the sequence $\{_h\theta\}$ has but a single limit point θ^*, which is in U', bounded. It follows that the sequence $\{_h\theta\}$ is convergent to θ^* and the sequence $\{_hf_{qr}\}$ is convergent to f_{qr}^*. The result of Equations 205 states that θ^*, the limit of the sequence $\{_h\theta\}$, satisfies the maximum likelihood equations for incomplete data, $\theta^* = \hat{\theta}$.

LITERATURE CITED

Clark, C. E. and Williams, G. T. 1960. Estimates from censored samples. Annals of the Institute of Mathematical Statistics, Tokyo, 12:209-220.

Cohen, A. C., Jr. 1949. On estimating the mean and standard deviation of truncated normal distributions. Journal of the American Statistical Association 44:518-525.

________. 1950. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. Annals of Mathematical Statistics 21:557-569.

________. 1954. Estimation of the Poisson parameter from truncated samples and from censored samples. Journal of the American Statistical Association 49:158-168.

Cramer, H. 1946. Mathematical methods of statistics. Princeton, N. J., Princeton University Press.

David, F. N. and Johnson, N. L. 1952. The truncated Poisson. Biometrics 8:275-285.

Deemer, W. L., Jr. and Votaw, D. F., Jr. 1955. Estimation of parameters of truncated or censored exponential distributions. Annals of Mathematical Statistics 26:498-504.

den Broeder, G. G. 1955. On parameter estimation for truncated Pearson Type III distributions. Annals of Mathematical Statistics 26:659-663.

Des Raj 1953. Estimation of the parameters of Type III populations from truncated samples. Journal of the American Statistical Association 48:336-349.

Finney, D. J. 1949. The truncated binomial distribution. Annals of Eugenics 14:319-328.

Fisher, R. A. 1931. Properties and applications of Hh functions. In Mathematical tables. Vol. 1. pp. xxvi-xxxv. London, British Association for the Advancement of Science.

Gjeddebaek, N. F. 1949. Contribution to the study of grouped observations I. Application of the method of maximum likelihood in case of normally distributed observations. Skandinavisk Aktuarietidskrift 32:135-159.

Gupta, A. K. 1952. Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika 39:260-273.

Hald, A. 1949. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. Skandinavisk Aktuarietidskrift 32:119-134.

Hartley, H. O. 1958. Maximum likelihood estimation from incomplete data. Biometrics 14:174-194.

Kendall, M. G. and Stuart, A. 1958. The advanced theory of statistics. Vol. 1. New York, N. Y., Hafner Publishing Company.

Krane, S. A. 1957. Maximum likelihood estimation from incomplete data for continuous distributions. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Moore, P. G. 1952. The estimation of the Poisson parameter from a truncated distribution. Biometrika 39:247-251.

________. 1954. A note on truncated Poisson distributions. Biometrics 10:402-406.

Pearson, K. and Lee, A. 1908. On the generalized probable error in multiple normal correlation. Biometrika 6:59-68.

Rider, P. R. 1953. Truncated Poisson distribution. Journal of the American Statistical Association 48:826-830.

________. 1955. Truncated binomial and negative binomial distributions. Journal of the American Statistical Association 50:877-883.

Sampford, M. R. 1955. The truncated negative binomial distribution. Biometrika 42:58-69.

Sarhan, A. E. and Greenberg, B. G. 1956. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. I. The normal distribution up to samples of size 10. Annals of Mathematical Statistics 27:427-451.

________ and ________. 1957. Tables for best linear estimates by order statistics of the parameters of single exponential distributions from singly and doubly censored samples. Journal of the American Statistical Association 52:58-87.

Tate, R. F. and Goen, R. L. 1958. Minimum variance unbiased estimation for the truncated Poisson distribution. Annals of Mathematical Statistics 29:755-765.

Tippett, L. H. C. 1932. A modified method of counting particles. Proceedings of the Royal Society of London, Series A, 137:434-446.

ACKNOWLEDGEMENTS

The author would like to express his appreciation to Professor H. O. Hartley for suggesting the problem of this thesis and sharing with the author some of his ideas on its solution, to Professor T. A. Bancroft, head of the Department of Statistics at Iowa State University of Science and Technology, for making the author's stay at the university possible and academically profitable, to the National Science Foundation and other agencies of the U. S. Government for financial support, and to Mrs. Barbara Konopik for composing and typing this copy of the thesis.